Data confidentiality and data security

United Nations | DESA Statistics Division

IOM UN MIGRATION

Data confidentiality

- The protection of data, which can be tracked to individual persons and businesses/institutions

- Need to protect the identity of persons and businesses/institutions so it is not:
    - Misused
    - Disclosed without authorization

Data confidentiality

Statistical confidentiality and privacy are set out in clear terms in Principle 6 of the United Nations Fundamental Principles of National Official Statistics (UNFPOS):

Principle 6 - Confidentiality

Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

UNSD- Fundamental principles

Data confidentiality and security increase trust in society

Protecting confidentiality and ensuring security help increase trust and goodwill from the persons and organisations whose data we are dependent on

United Nations | DESA Statistics Division

IOM UN MIGRATION

What are personal data?

Article 4 EU GDPR: “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

National laws on data protection

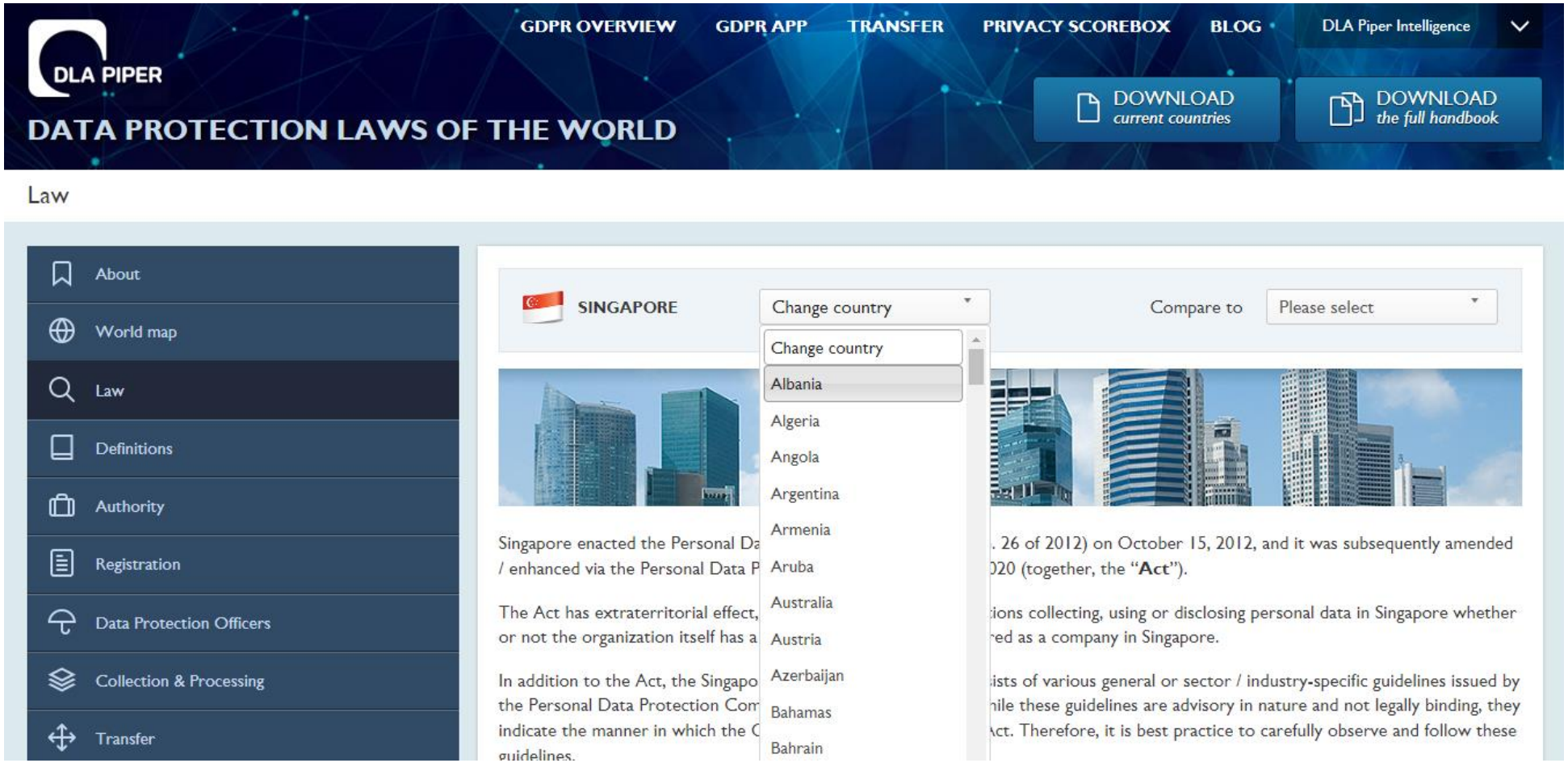
Protection of personal information might be granted in national laws e.g.:

- European Union:
  - General Data Protection Regulation (GDPR)

The processing of personal data for statistical purposes is permitted if there is a legitimate basis and appropriate safeguards to protect individuals' rights and privacy, by ensuring that individuals cannot be identified, whether directly or indirectly.

- Namibia:
  - Namibia Statistics Act 35 9d
  - Draft data protection bill

Every employee of the government… engaged in.. collection or to whom such statistics or information is made available, must take an oath of confidentiality



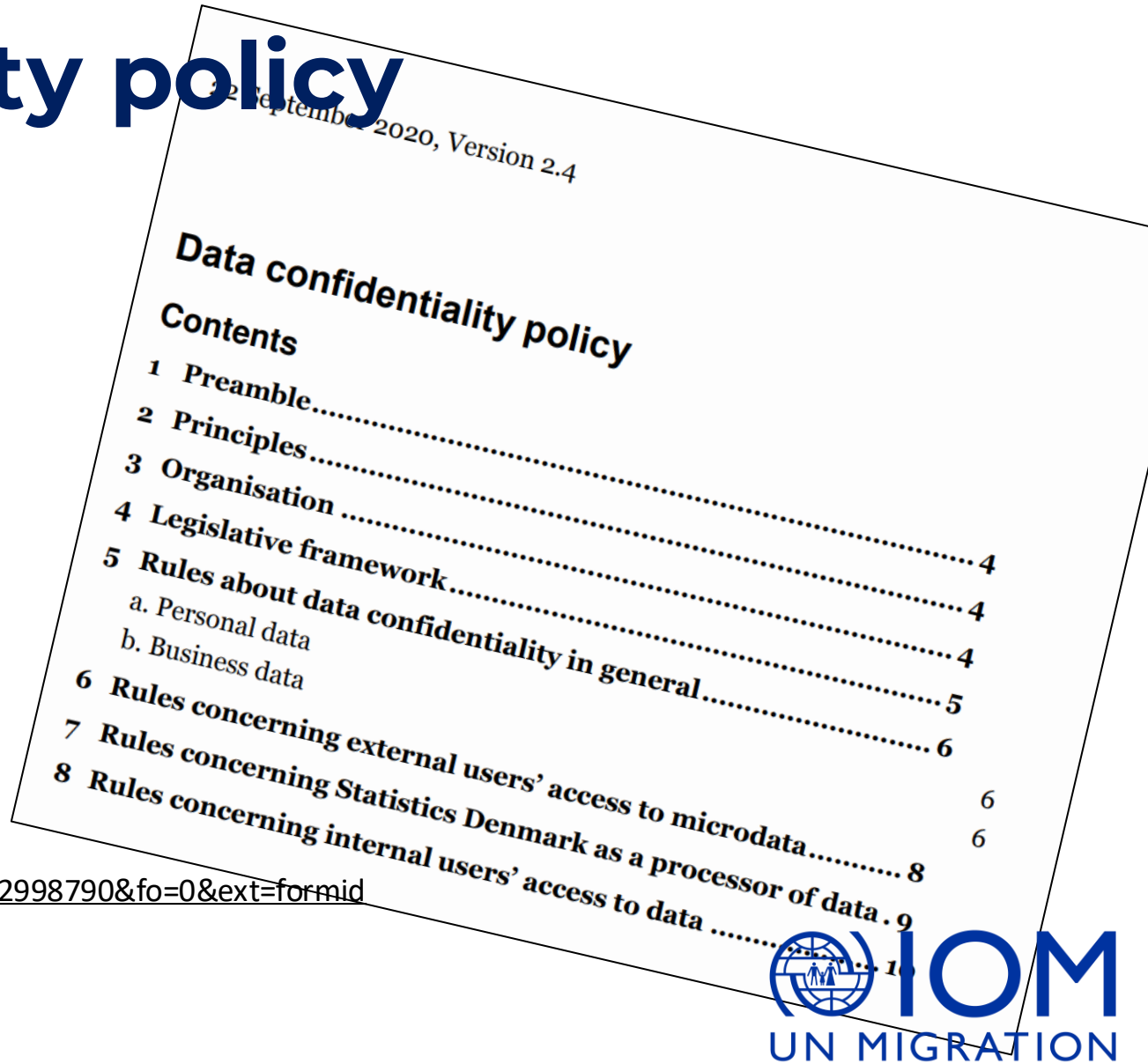
https://www.dlapiperdataprotection.com/index.html?c=SG&t=law#

United Nations | DESA Statistics Division

IOM UN MIGRATION

Data confidentiality policy

- Some countries have decided to publish a policy on data confidentiality
- This may be a good idea to ensure transparency on the approach taken



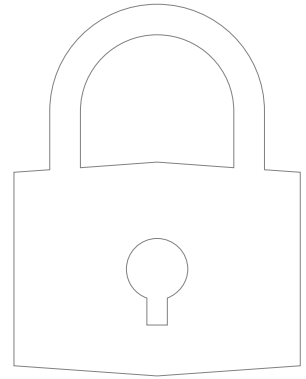
2 September 2020, Version 2.4

Data confidentiality policy

Contents

Source: https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=502998790&fo=0&ext=formid

How do you protect confidential data?



There are different mechanisms you can put in place to protect data e.g.

- Raise awareness among staff on confidentiality and why it is important
- Ensure staff are committed to safeguard confidentiality even after their work is finished or they leave the institution
- Restrict access to confidential data so only those who need to have access, have access
- Set up procedures for granting access incl. management approval for getting access
- Confidential data is stored in dedicated databases
- Log access to confidential data so the institutions can see who looks at the data
- Practical measures like locking USB access at work computers

Data security

- How data is protected from external and internal risks e.g.
  - Misuse
  - Involuntary deletion of data
  - Unauthorized access
  - Limitation of access
  - Loss of data

How do you secure data from external risks

Measure to reduce external risks e.g.:

- Securing the internal network from hackers and malware by setting up firewalls, antivirus systems and email-protection system (SPAM control, malicious code)
- Putting in place back-up procedures e.g., to copy data at a decided frequency (daily, weekly, monthly as needed)
- Setting up a physical “disaster site” where all data is transferred on a regular basis e.g., in case of fire in the server room which will eradicate all data – therefore always reduce fire risks in the server room!

How do you secure data from internal risks

Measure to reduce internal risks e.g.:

- Raising awareness of staff on how to improve data security
- Storing data on a centralised server (databases) prevent random storage of confidential information e.g., on someone's hard desk + possibility to decide on security measures which can be centrally managed
- Register (log) which employees access data

Staff awareness
Example: Denmark

”8 rules on good staff behavior

1. Your password is strictly personal
2. Activate your screen saver when you leave your PC
3. Delete emails that you do not trust - without reading it
4. Confidential data should be stored on dedicated network shares
5. Do not save documents on your PC’s hard desk
6. Take special care when working with confidential data outside Statistics Denmark
7. Encourage data providers to deliver confidential data using encrypted communication lines
8. Contact IT Division if you have questions about IT security”

Source: Statistics Denmark

Identifiable information

Direct identifiers: Any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records;

Indirect identifiers: Any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information

Concrete examples

Name	Date of birth	Gender	Age	Citizenship	City of birth	Country of residence	City of residence
-expunged-	-expunged-	F	34	Slovenia	Koper	France	Annecy

- Can you identify the individual? Under what kinds of scenarios?
- Risk = probability of disclosure x impact
- (Legitimate) Purpose and proportionality

Concrete examples

Name	Date of birth	Gender	Age	Citizenship	City of birth	Country of residence	City of residence
-expunged-	-expunged-	F	34	Slovenie	Koper	France	Annecy

Gender	Citizenship	Occupation	Age	Country of birth	Number of cases in the database
F	-	-	-	-	534
F	Canadian	-	-	-	178
F	Canadian	Engineer	-	-	32
F	Canadian	Engineer	48	-	4
F	Canadian	Engineer	48	Nigeria	1

Example

Table (i) Dataset

Birth date	Gender	Post code
13/01/90	M	33710
25/10/90	M	33410
18/08/89	F	55810
14/02/82	F	94320
30/11/82	F	94870

Example: Reduction in details

Table (i) Dataset

Birth date	Gender	Post code
13/01/90	M	33710
25/10/90	M	33410
18/08/89	F	55810
14/02/82	F	94320
30/11/82	F	94870

Table (ii) Dataset – reduction in details

Birth year	Gender	Post code
90	M	33***
90	M	33***
89	F	55***
82	F	94***
82	F	94***

Example: K-anonymization

Table (i) Dataset

Birth date	Gender	Post code
13/01/90	M	33710
25/10/90	M	33410
18/08/89	F	55810
14/02/82	F	94320
30/11/82	F	94870

Table (iii) Dataset – reduction in details + k-anonymization (k=2)

Birth year	Gender	Post code
90	M	33***
90	M	33***
REMOVED	*REMOVED*	*REMOVED*
82	F	94***
82	F	94***

Example: Synthetic data

Start with an empty dataset of the same structure

- For each attribute, sample one attribute value
- After sampling one attribute value for each attribute, you have a candidate synthetic record
- Check if the synthetic record belongs to the safe or risky subset defined (k=e.g. 2)
- If safe, keep, if risky, discard

Table (ii) Dataset – reduction in details

Birth year	Gender	Post code
90	M	33***
90	M	33***
89	F	55***
82	F	94***
82	F	94***

Example: Synthetic data

Start with an empty dataset of the same structure

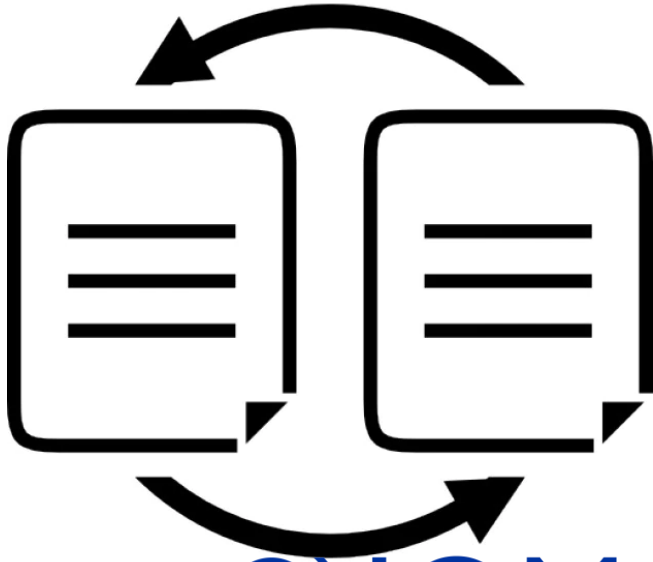
- For each attribute, sample one attribute value
- After sampling one attribute value for each attribute, you have a candidate synthetic record
- Check if the synthetic record belongs to the safe or risky subset defined (k=e.g. 2)
- If safe, keep, if risky, discard

Table (ii) Dataset – reduction in details

Birth year	Gender	Post code
90	M	33***
90	M	33***
89	F	55***
82	F	94***
82	F	94***

Secure data transmission

- Effective data transmission provisions seek to ensure that data is transmitted in a secure way, so data is not leaked nor lost.
    - Data can for example be encrypted

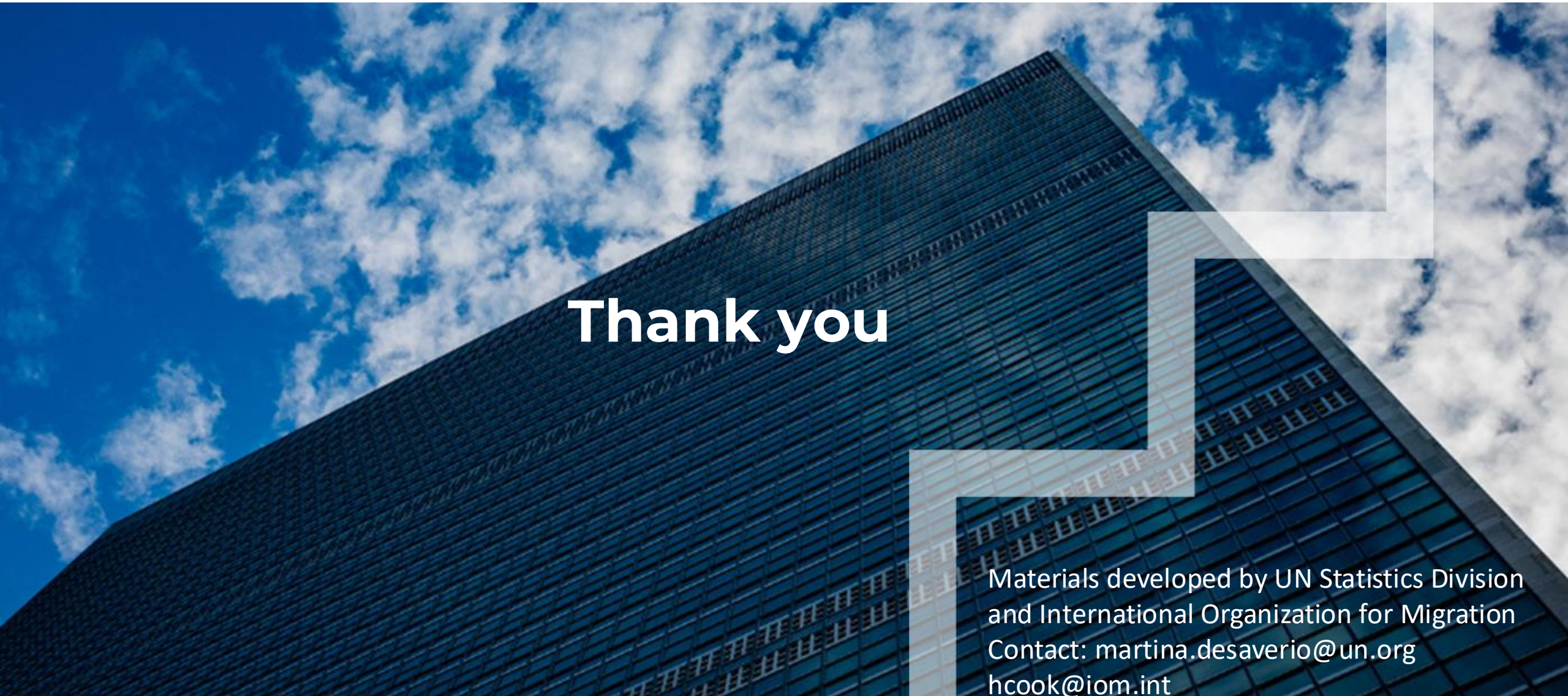


Resources

- The **UN Guide on Privacy-enhancing technologies for official statistics** is a great starting point: https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf
- Making each case count: Leveraging Administrative Data on Trafficking in Persons
- Algorithm public and open source: https://github.com/microsoft/synthetic-data-showcase
- Example of a synthetic database using this algorithm: https://www.ctdatacollaborative.org/global-synthetic-dataset
- Handbook on Management and Organization of National Statistical Systems

United Nations | DESA Statistics Division

IOM UN MIGRATION



Thank you

Materials developed by UN Statistics Division and International Organization for Migration

Contact: martina.desaverio@un.org
hcook@iom.int