9-26-2023

Data Ethics and Privacy for Researchers

Kelley F. Rowan
Florida International University, krowan@fiu.edu

Data ethics and privacy for researchers

Kelley Rowan, Digital Archives Librarian

Agenda

Laws, Torts, and Regulations

Definitions, Getting Started, and Workflows

PII v. Personal (Sensitive) Data

Anonymization Techniques & Tools

Encryption Resources

Avoiding Anonymization

Data Anonymization Services

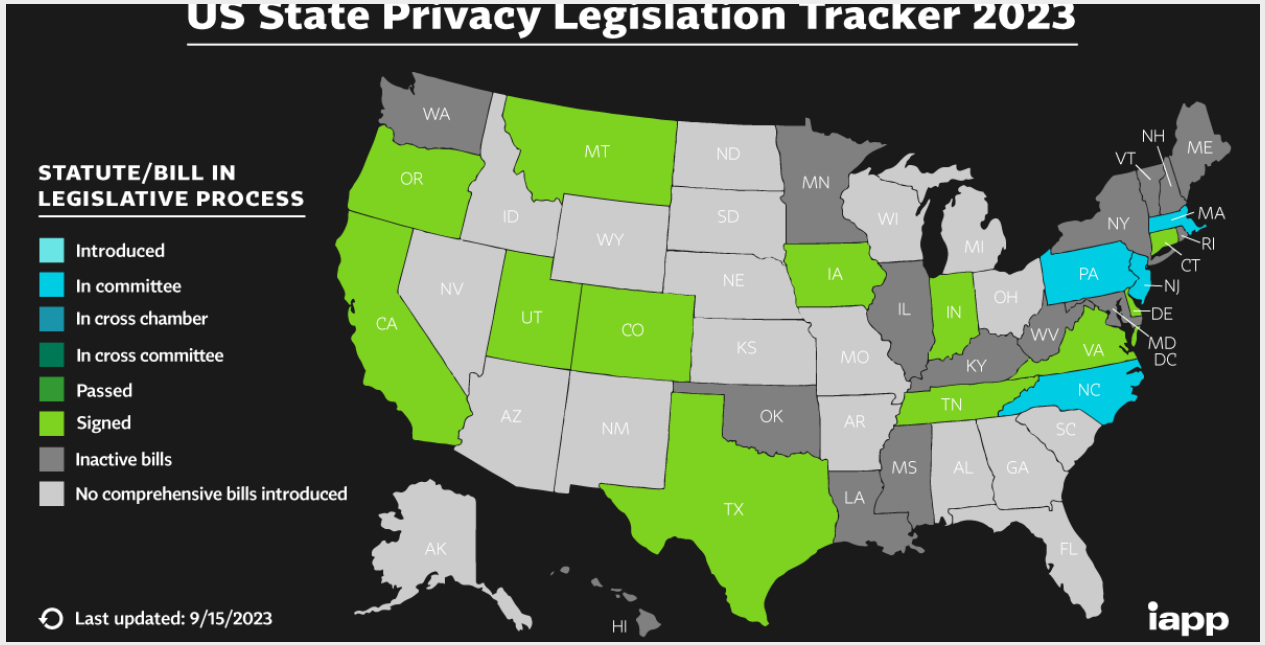
US Federal regulations

- **Privacy Act of 1974**
  - Governs federal records
  - https://www.justice.gov/opcl/privacy-act-1974
- **Children's Online Privacy Protection Act of 1988, 15 U.S.C 6501-6505**
  - https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule
- **Gramm-Leach Bliley Act, 1999 (Financial Modernization Act)**
  - Governs banks and private financial information
  - https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act
- **Family Educational Rights and Privacy Act (FERPA)**
  - **https://studentprivacy.ed.gov/ferpa**
- **Health Insurance Portability & Accountability Act (HIPAA)**
  - https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

4 Privacy Torts: Intrusion into Seclusion; Public disclosure of private facts and data

State Privacy Laws

https://iapp.org/resources/topics/us-state-privacy/



TITLE 1.81.5. California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100]

CCPA & CPRA

link: https://oag.ca.gov/privacy/ccpa

California Privacy Rights Act – 2023

Link to excellent summary:

https://www.customerlabs.com/blog/ccpa-cpra-unraveling-the-mystery-in-2023/

Fully enforced beginning July 1, 2023

International Law

GDPR

General Data Protection Regulation (2018)

https://gdpr.eu/

*Secure people's data

*Make it easy for people to exercise control over their data

- **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
- **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
- **Data minimization** — You should collect and process only as much data as absolutely necessary for the purposes specified.
- **Accuracy** — You must keep personal data accurate and up to date.
- **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.
- **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
- **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

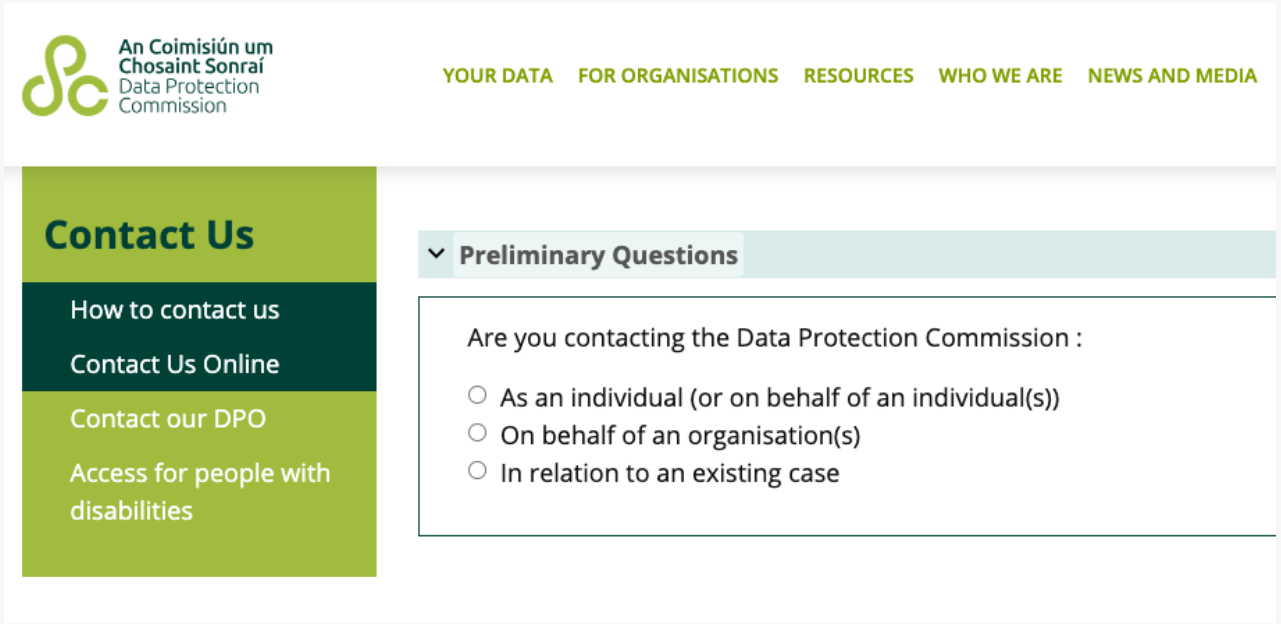
RTBF

Right to be Forgotten
- erasure obligations

RTBI

Right to be Informed
- must provide a privacy policy

GDPR non-compliance consequences



$21,150,500 or 4% of profits, whichever is greater

Enforcement tracker:

https://www.enforcementtracker.com/

Chapter 1, Article 4 (GDPR) makes no distinction between companies and individuals

- **PCI DSS**

Payment Card Industry Data Security Standards

https://www.pcisecuritystandards.org/

PCI Data Handling Standards

- Do not keep physical copies
- NEVER store the CVV
- Store in a secure locked place
- Limit personnel with access
- Use security cameras in area
- Assess and securely destroy every 3 months
- Provide a privacy plan for clients

Definitions

GDPR Article 4(1)

“personal data” pertains to “any information relating to an identified or identifiable natural person (‘data subject’)"

PII (Personally Identifiable Information) = information that identifies an individual

Sensitive Information = information that may create harmf if exposed

Data subject = person

Before collecting data…

Determine your risks (impact assessement)

Develop a security plan (VPN, data anonymization)

Share your privacy policy

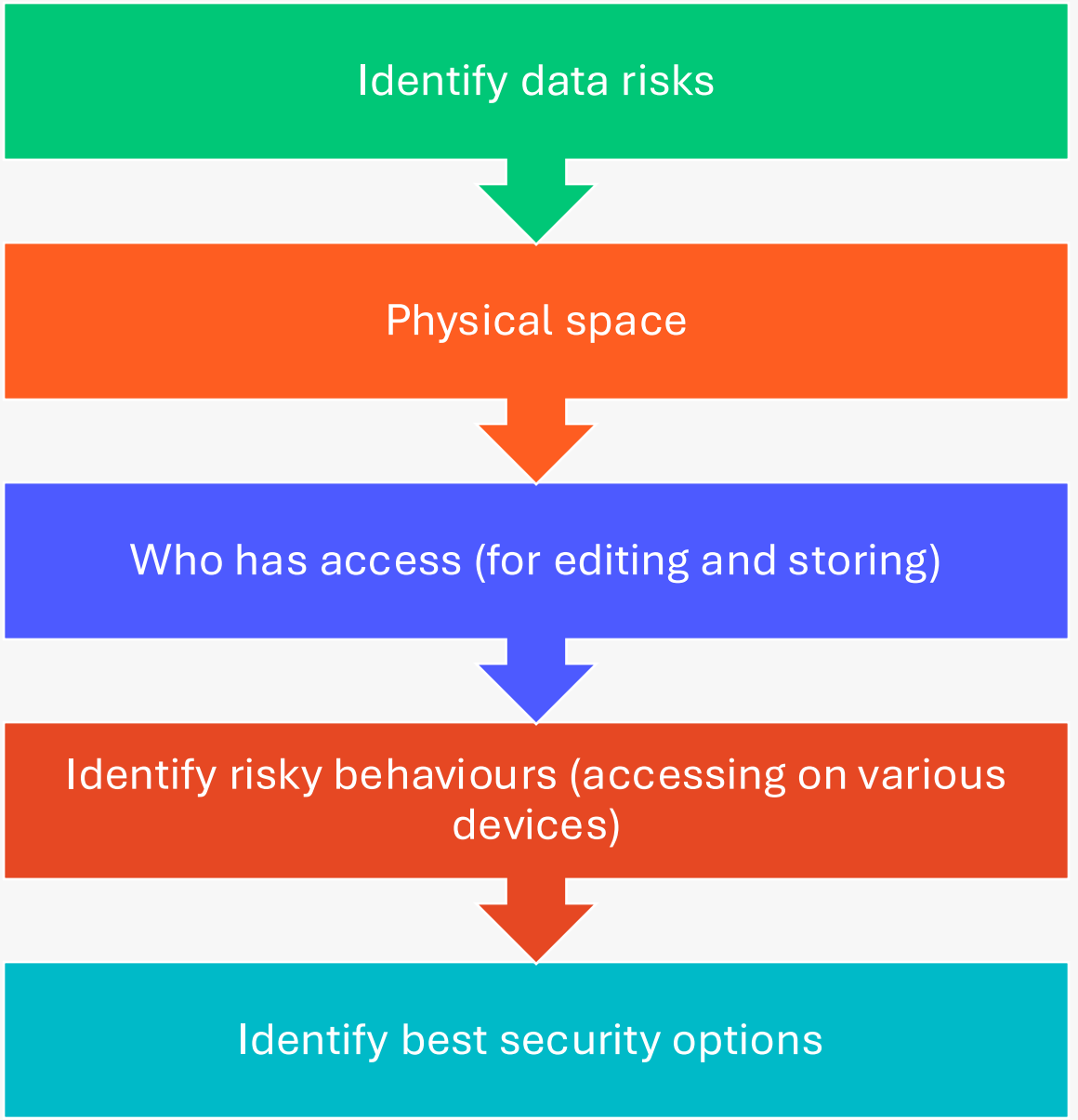
Impact Assessment

An impact assessment must always be conducted in high-risk situations such as when collecting, storing, and/or processing personal (sensitive) data.

- **What are the consequences of a breach?**
- **How will you secure personal data?**

(encryption, VPN, de-identification, key storage, 3rd party encryption, data storage and destruction)

Impact assessment



Privacy Policy

- ✓ *What data will be collected*
- ✓ *How it is collected*
- ✓ *How it will be used and for what purpose*
- ✓ *Who has access*
- ✓ *How the data is stored and for how long*
- ✓ *How the data will be securely destroyed*
- ✓ *Contact information*

Keeping data safe with anonymization techniques

Identify

Identify your PII and sensitive data

Ascertain

Ascertain whether other individuals will need working access to the data

Determine

Determine whether the data will be published

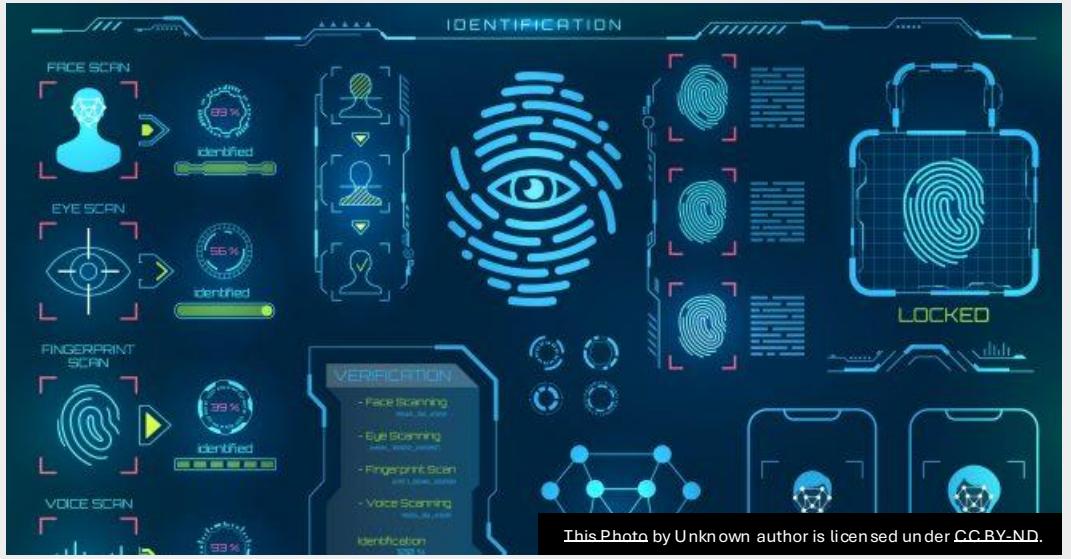
Develop a key

Secure your key

Computer Security

- ✓ *Use passphrases with 2FA*
- ✓ *Use biometrics with 2FA*
- ✓ *Use GDPR compliant tools*

https://gdpr.eu/compliant-services/



Workflow for teams with access to personal data

Ideal:

one person has access and anonymizes before granting access to others

Good:

2+ trusted people may work on collecting and anonymizing, including secure storage of a key.

Weak:

Everyone in the workplace has access, anonymization happens at the end before publishing.

Identifying PII and personal data

	A	B	C	D	E	F	G	H	I	J	K
1	Last Name	First Name	FI #	DC# ingest	DC# published	Notes	published	Full text pdf/a	embargo	M/PhD	Major
2	Sale	Tonia	FIDC000332	8431	3670		12.10.1998	12.10.1998		MS	Hospitality
3	Olinger	Patricia	FIDC000333	8432	3671		12.10.1998	12.10.1998	2 yrs.	PhD	Business Admin
4	Bennett	William	FIDC000334	8433	3672		12.10.1998	12.10.1998	2 yrs.	MS	speech pathology
5	Graham	Sharon	FIDC000335	8434	3673		12.10.1998	12.10.1998	1 yr.	PhD	GSS
6	Nelson	Thomas	FIDC000336	8435	3674		12.10.1998	12.10.1998		PhD	biomedical engineering
7	Warren	Carmen	FIDC000337	8436	3675		12.10.1998	12.10.1998		MS	electrical engineering
8	Scalf	Raymond	FIDC000338	8437	3676		12.10.1998	12.10.1998		MS	computer engineering

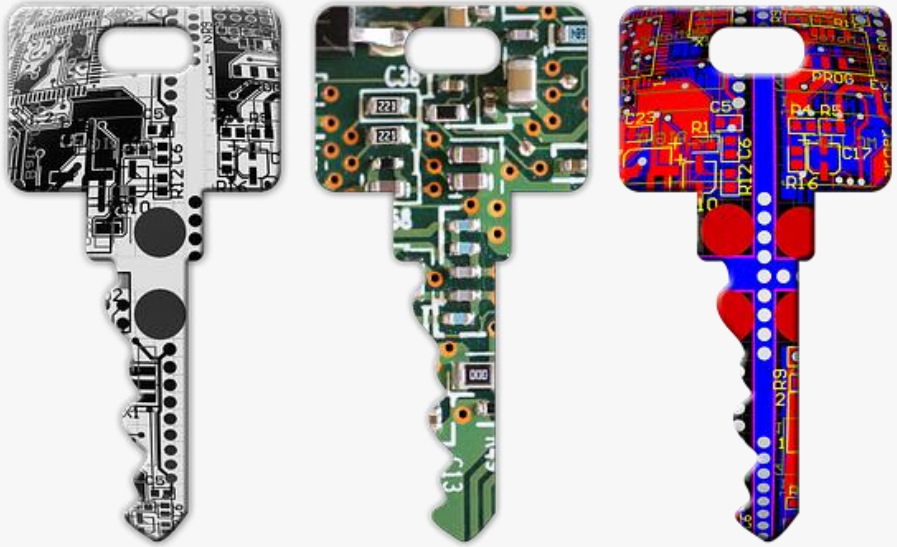
Look for possible points of re-identification!

Pseudonymization (masking)

The processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.

Use multiple pseudonymization techniques to achieve anonymization.

Secure the key.



Anonymized data

Anonymization is the irreversible removal of information that could lead to an individual being identified.

No longer considered PII by the GDPR.

You can achieve anonymization by combining various pseudonymization techniques.

Substitution

	A	B	C
1	Last Name	First Name	FI #
2	Sale	Tonia	FIDC000332
3	Olinger	Patricia	FIDC000333
4	Bennett	William	FIDC000334
5	Graham	Sharon	FIDC000335
6	Nelson	Thomas	FIDC000336
7	Warren	Carmen	FIDC000337
8	Scalf	Raymond	FIDC000338
9	Fossum	Stephen	FIDC000339
10	Stevens	Aubrey	FIDC000340
11	Pierce	Nicole	FIDC000341

Definition:

Replacing one column or row of data with completely different values. These could be names or numbers.

Effectiveness:

Highly effective in masking data and securing privacy. Pseudonymization.

Considerations:

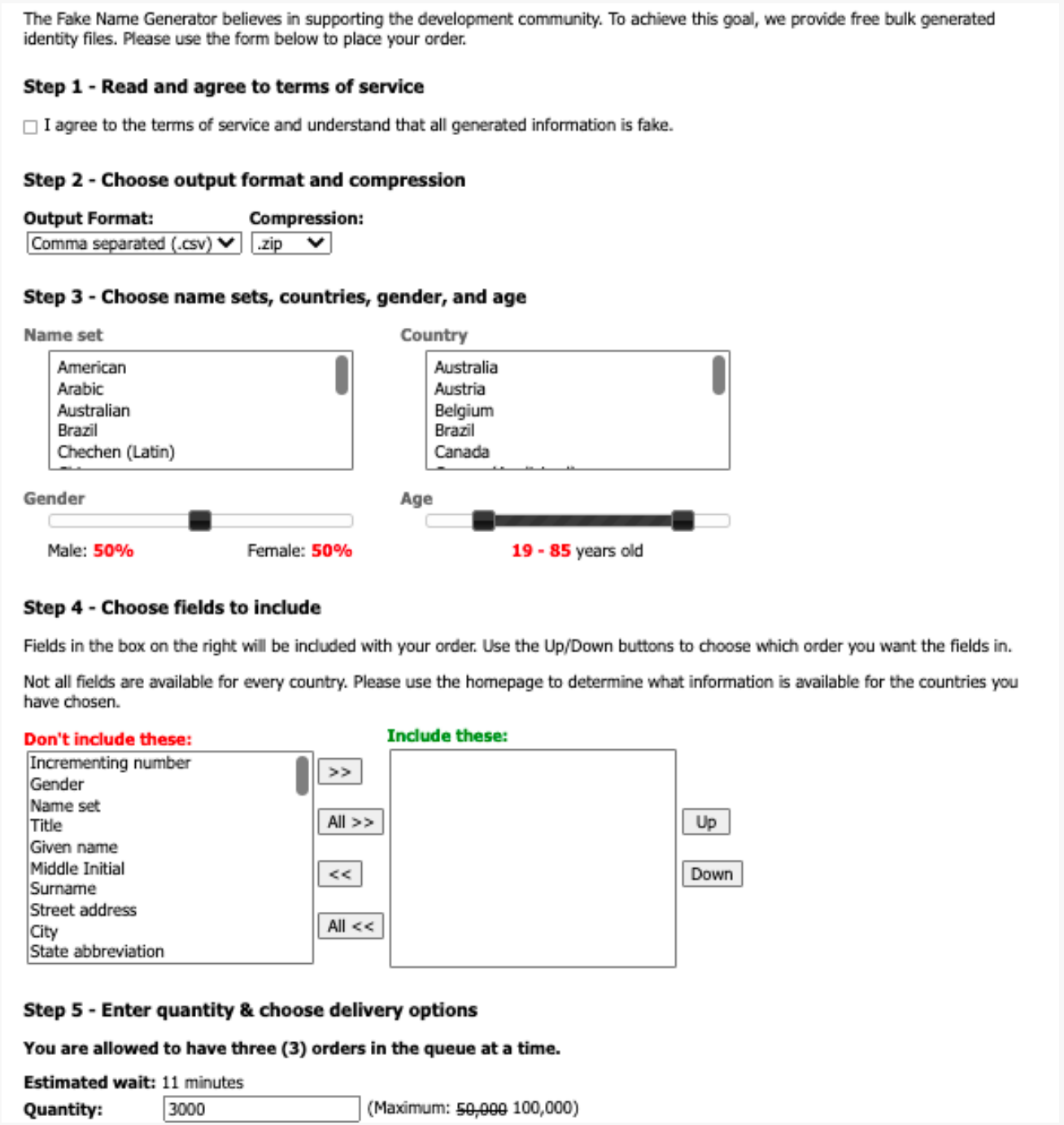
Full substitution of an entire group of data can attain full anonymization status and will no longer be considered sensitive data by the GDPR.

Tools:

fake name generator (free)

https://www.fakenamegenerator.com/order.php

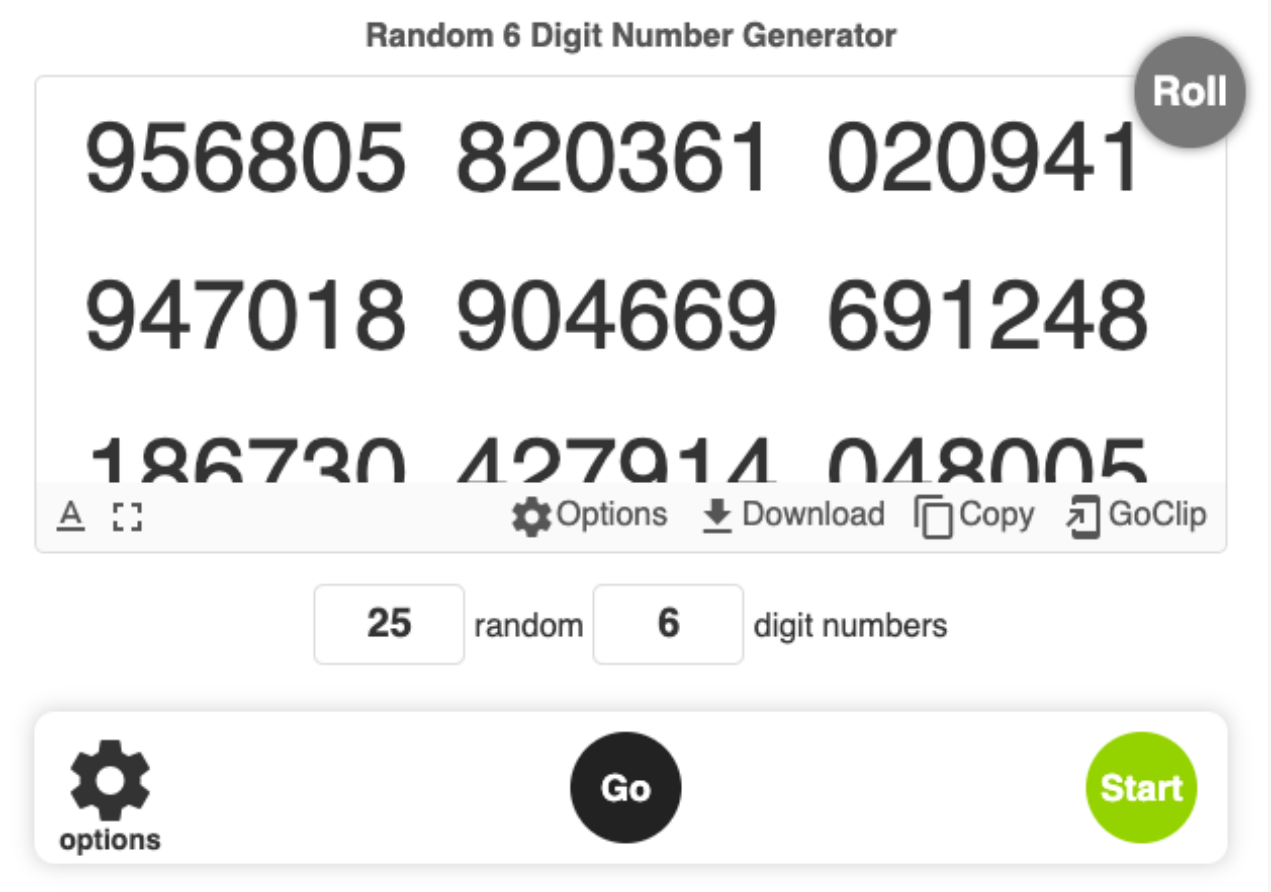
	A	B	C
1	GivenName	Surname	
2	Tonia	Sale	
3	Patricia	Olinger	
4	William	Bennett	
5	Sharon	Graham	
6	Thomas	Nelson	
7	Carmen	Warren	
8	Raymond	Scalf	
9	Stephen	Fossum	
10	Aubrey	Stevens	
11	Nicole	Pierce	
12	Lillian	Saenz	



Tools:

Random number generator

https://numbergenerator.org/random-6-digit-number-generator#!numbers=25&length=6&addfilters=



Nulling out

Last Name	First Name	FI #	DC# ingest	DC# published	Notes
XXXX	XXXX	FIDC000350	8449	3688	
XXXX	XXXX	FIDC000351	8450	3689	
XXXX	XXXX	FIDC000352	8451	3690	
XXXX	XXXX	FIDC000353	8452	3691	
XXXX	XXXX	FIDC000354	8453	3692	

Definition:

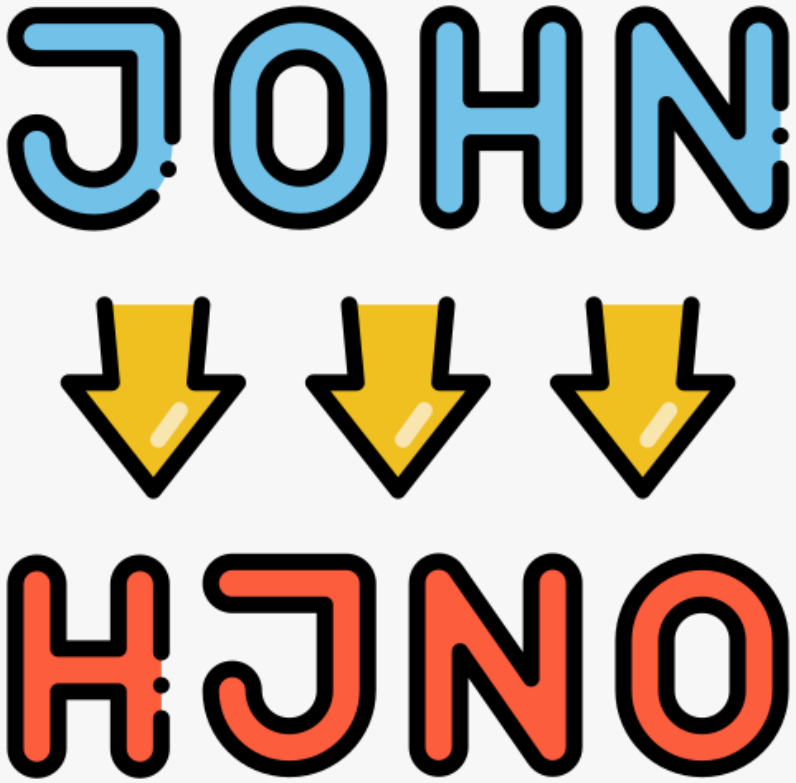
Replacing a data field with a null value.

Effectiveness:

Highly effective. Achieves anonymization.

Considerations:

Often reduces data integrity.



Scrambling

Definition:

Scrambling the letters or numbers in a data field.

Effectiveness:

Weak form of pseudonymization. The data is susceptible to being "unscrambled" and re-identified. Can be stronger for long number sequences where the same scrambling algorithm is not used in each data field.

Tools:

word scrambler

https://www.wordunscrambler.net/word-scrambler.aspx

First Name	FI #	DC# ingest	DC# published
Tonia	FIDC000332	8431	3670
Patricia	FIDC000333	8432	3671
William	FIDC000334	8433	3672
Sharon	FIDC000335	8434	3673
Thomas	FIDC000336	8435	3674
Carmen	FIDC000337	8436	3675
Raymond	FIDC000338	8437	3676
Stephen	FIDC000339	8438	3677

First Name	FI #	DC# ingest	DC# published
Tonia	FIDC000332	8431	3670
Stephen	FIDC000333	8432	3671
Raymond	FIDC000334	8433	3672
Carmen	FIDC000335	8434	3673
Thomas	FIDC000336	8435	3674
Sharon	FIDC000337	8436	3675
William	FIDC000338	8437	3676
Patricia	FIDC000339	8438	3677

Shuffling

Definition:

Shuffling the values in a data field.

Effectiveness:

Weak form of pseudonymization if used alone. Susceptible to re-identification by determining the shuffling algorithm. Can achieve anonymization when used with other masking techniques.

published	Full text pdf/a
12.11.1998	3.16.1996
12.11.1998	3.16.1996
12.11.1998	3.16.1996
12.12.1998	3.16.1996
12.12.1998	3.16.1996

Date Aging

Definition:

Choosing a random number of days to "age" a date.

Effectiveness:

Intermediate to strong form of pseudonymization. Somewhat susceptible to re-identification by determining the aging value.

Variance

Definition:

varying the date and number values. Common usage is with financial data.

Example:

for number values +/-10%; dates +/- 200 days

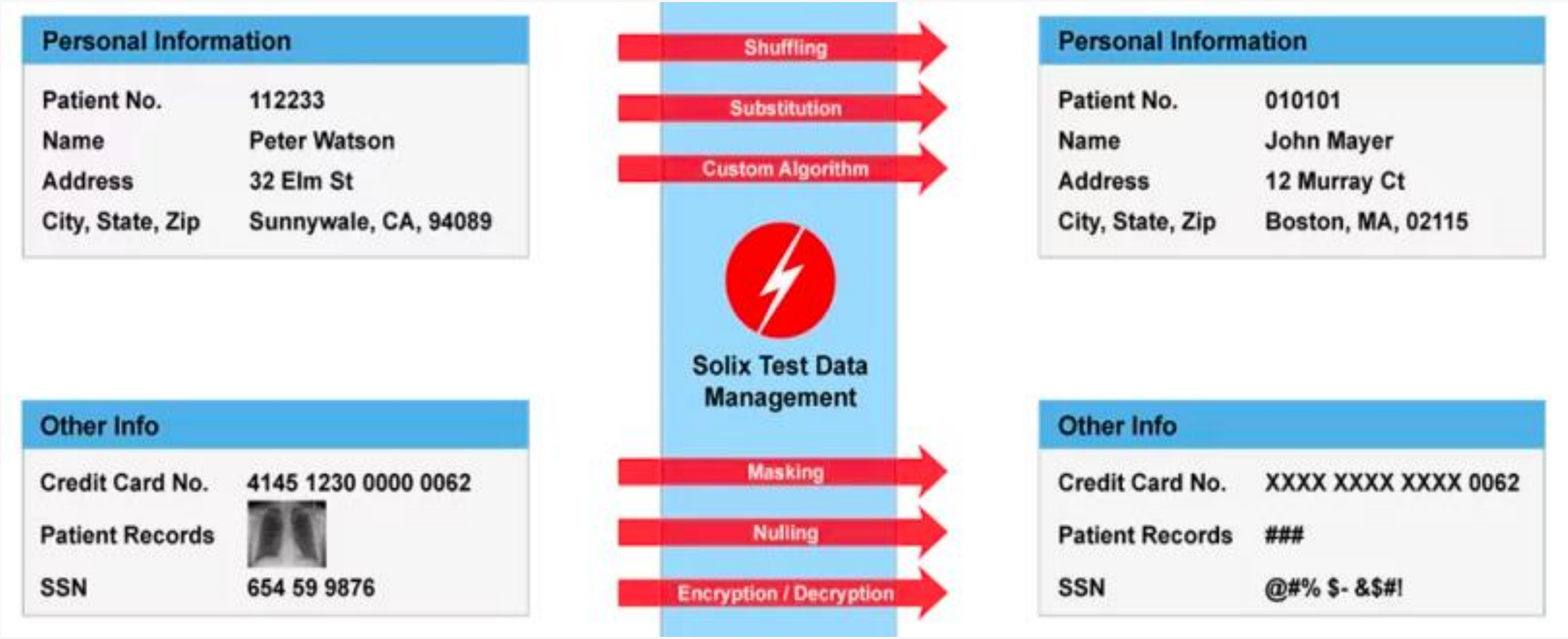
Masking out

Definition:

Hiding some (not all) of the digits in a field.

e.g. xxxx-xxxx-xxxx-1079

Examples


Personal Information
Patient No. 112233
Name Peter Watson
Address 32 Elm St
City, State, Zip Sunnywale, CA, 94089
Other Info
Credit Card No. 4145 1230 0000 0062
Patient Records
SSN 654 59 9876
Shuffling
Substitution
Custom Algorithm
Solix Test Data Management
Masking
Nulling
Encryption / Decryption
Personal Information
Patient No. 010101
Name John Mayer
Address 12 Murray Ct
City, State, Zip Boston, MA, 02115
Other Info
Credit Card No. XXXX XXXX XXXX 0062
Patient Records ###
SSN @#% $- &$#!

Size of the data group

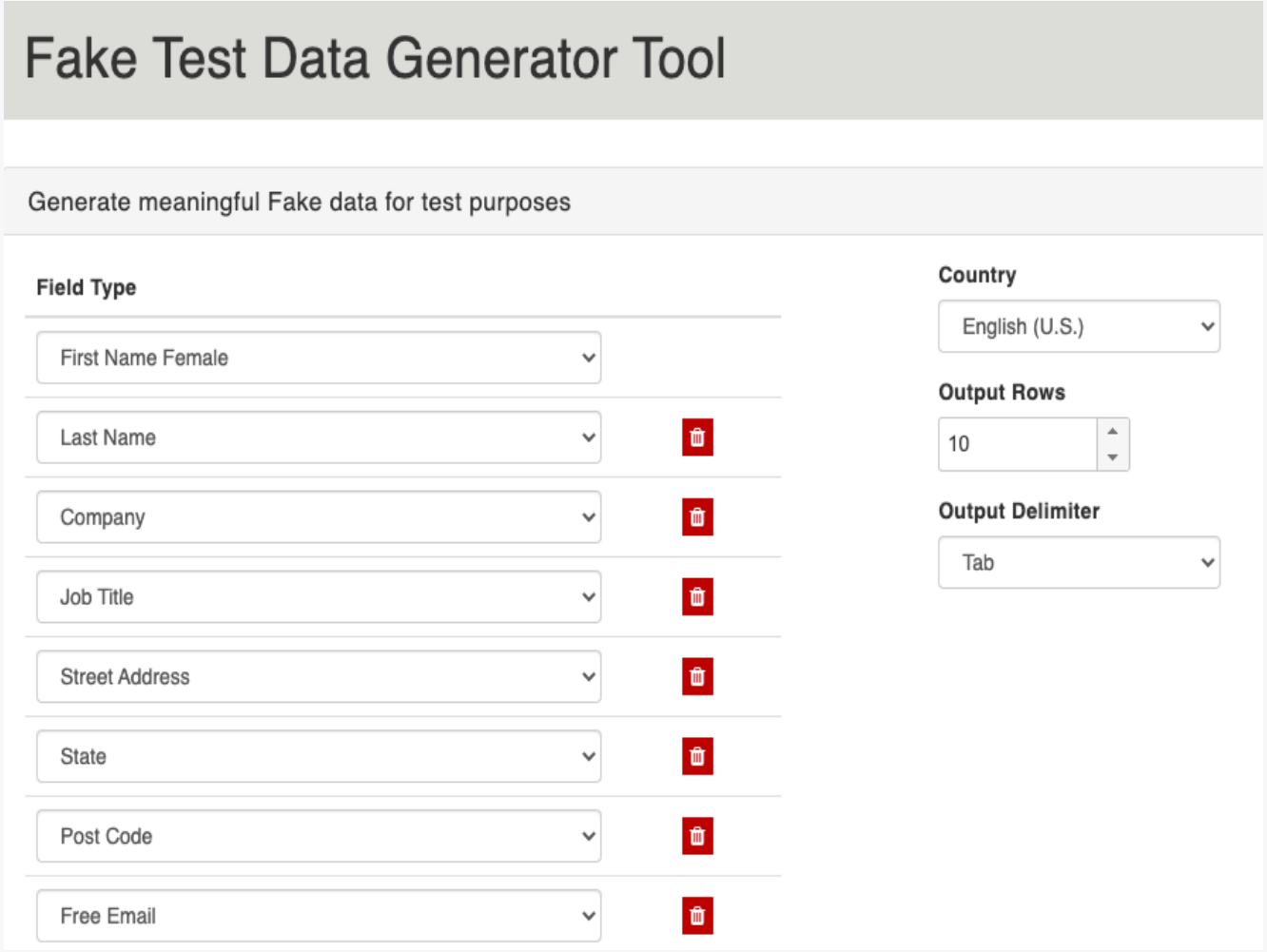
A small group of data subjects undermine most forms of pseudonymization as anyone with knowledge of the subjects can re-identify the data.

Tools:

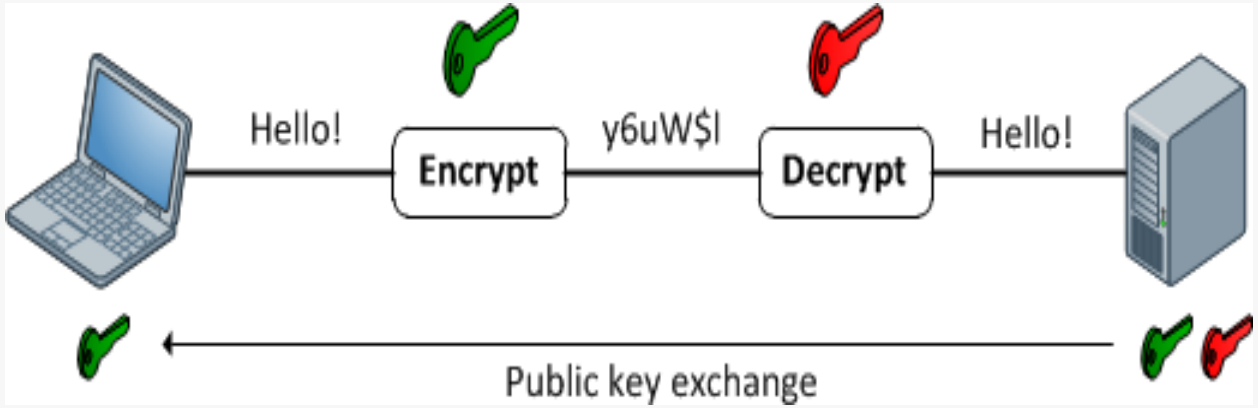
Fake data

https://www.coderstool.com/fake-test-data

Evelyn Feest Marvin Ltd Fitness Trainer 293 Vincenzo Parkway Louisiana 24060-9543 littel@gmail.com
Loraine Olson McLaughlin Group Agricultural Science Technician 13473 Brown Route Massachusetts 91737
turcotte.cielo@hotmail.com
Kiara Bashirian Schultz, Leannon and Miller Telemarketer 24494 Raleigh Plaza Apt. 132 West Virginia 14488 xwalker@hotmail.com
Ashley O'Keefe Lebsack-Schinner Drilling and Boring Machine Tool Setter 90882 Lynch Extensions Tennessee 59634
price52@hotmail.com
Chelsie Muller Metz, Rohan and Paccocha Shipping and Receiving Clerk 2338 Reichert Passage Iowa 63199
loraine.corwin@hotmail.com



Encryption



Definition:

An algorithm masks the data for you and requires a key to un-encrypt for editing and usage.

Effectiveness:

Highly effective unless the key or password to the encrypted folder is compromised.

Considerations:

Mobile versions can be less safe.

Free encryption options

1. Microsoft OneDrive
(personal vault not available for macOS)

2. Proton Drive, VPN,
Email: https://proton.me/drive

3. Folder Lock

https://www.newsoftwares.net/folderlock/

4. AxCrypt

https://www.axcrypt.net/pricing/

5. VeraCrypt

https://www.veracrypt.fr/en/Downloads.html

Other encryption options

https://www.techradar.com/best/best-encryption-software

For Windows

1. Secure IT 2000
2. SensiGuard
3. Renee File Protector

For macOS

1. Concealer

How to avoid anonymization

1

Do not collect any personal data

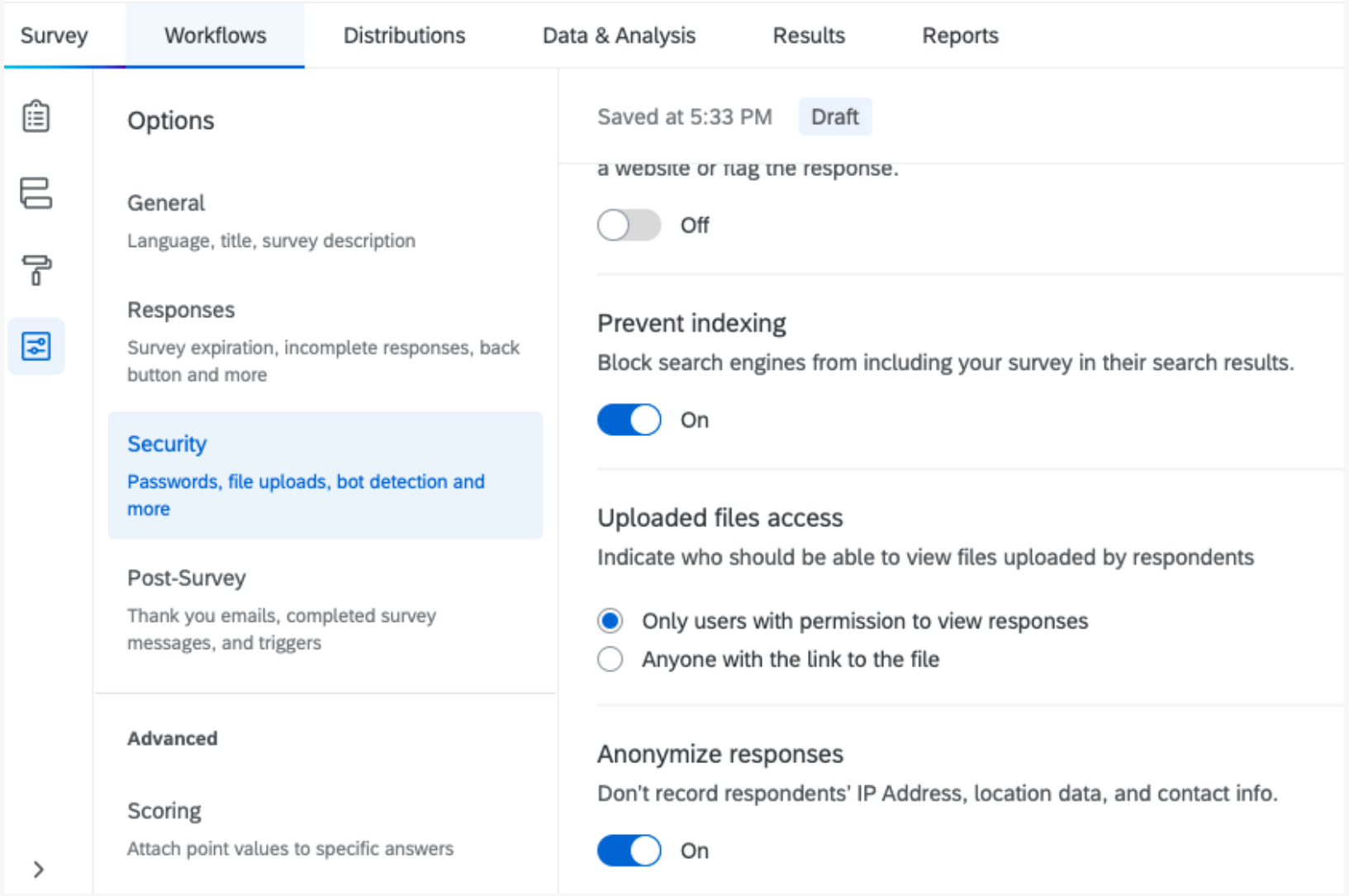
2

Do not store any personal data

3

Do not share any personal data

Qualtrics


Survey
Workflows
Distributions
Data & Analysis
Results
Reports
Options
General
Language, title, survey description
Responses
Survey expiration, incomplete responses, back button and more
Security
Passwords, file uploads, bot detection and more
Post-Survey
Thank you emails, completed survey messages, and triggers
Advanced
Scoring
Attach point values to specific answers
Saved at 5:33 PM
Draft
a website or flag the response.
Off
Prevent indexing
Block search engines from including your survey in their search results.
On
Uploaded files access
Indicate who should be able to view files uploaded by respondents
Only users with permission to view responses
Anyone with the link to the file
Anonymize responses
Don't record respondents' IP Address, location data, and contact info.
On

Data anonymization services

1. **Accelario**

https://accelario.com/

2. **Anonos**

https://www.anonos.com/product-overview?hsLang=en

3. **K2View**

https://www.k2view.com/

Considerations:

Be sure services used do not receive actual data, but data already encrypted.

Link to slides:
https://digitalcommons.fiu.edu/glworks/139/

Kelley Rowan

Digital Archives Librarian

krowan@fiu.edu