



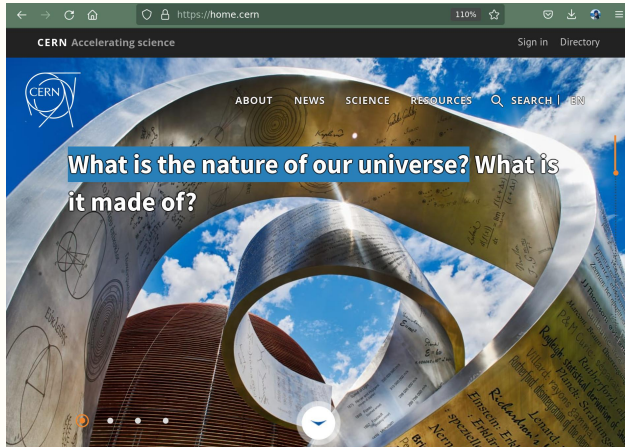
**Opendata.ch /
2024 Forum**

24.05.2024
Uni FR - Fribourg

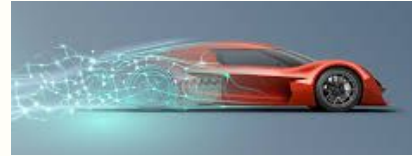
PRIVACY / OPENDATA THE HIDDEN COST

Konstantinos Samaras-Tsakiris
AI4PRIVACY ADVISOR

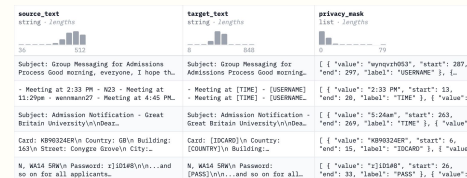
ABOUT ME



TECHNOLOGY



4 Ai4Privacy



OPENNESS

&

PRIVACY

:

FRIENDS

~

FOES?

a) Course of overall sum of fines (cumulative):



Compliance lag?
Priced in?

GDPR FINES

SOURCES: enforcementtracker.com
tradingview.com

4 EASY STEPS FOR GDPR COMPLIANCE



“raw data, now” 
2009



WWW
1989

FAIR data 2016

Findable
Accessible
Interoperable
Reusable



Your data,
your choice
2018



inrupt

enterprise-grade **Solid**
software and services
2021

standard, open, interoperable
data formats and protocols
→ TRUSTWORTHY

sustainable commercially

SOURCES: [The State of Open Data](#) (book)
solidproject.org
[Vanityfair: Tim Berners Lee on Solid](#)

OPEN DATA



Your data,
your choice

PRIVACY: MISSED OPPORTUNITY?

“Every data enterprise must be a responsible steward of data, algorithms, and AI.”

OUR LATEST INITIATIVE

Data Provenance Standards

The first cross-industry metadata to bring transparency to the origin of datasets used for both traditional data and AI applications.

61% of CEOs in the [2023 annual IBM Institute for Business Value CEO study](#) cite the lack of clarity on data lineage and provenance as a top barrier to adoption of generative AI.

STANDARD	METADATA
+ Lineage	
+ Source	
+ Legal rights	
×	Protected data classification
Privacy and protection	Applied privacy enhancing techniques
+ Generation date	
+ Data type	
+ Generation method	
+ Intended use and restrictions	

Privacy and protection

Identifies any types of sensitive data associated with the current dataset and any privacy enhancing techniques applied

PRIVACY: STANDARD

SOURCES: dataandtrustalliance.org
[data&trust blog](https://data&trust.blog)

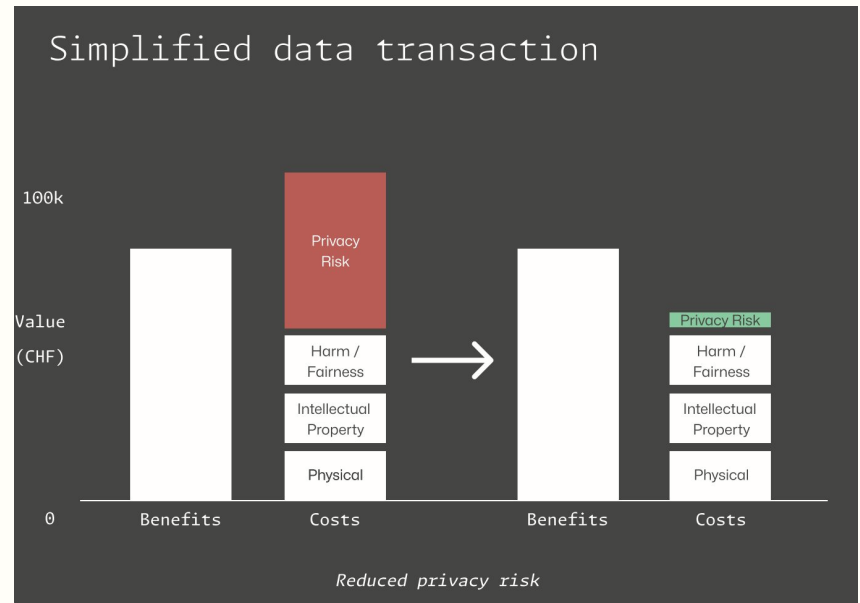
PRIVACY IS PROVENANCE

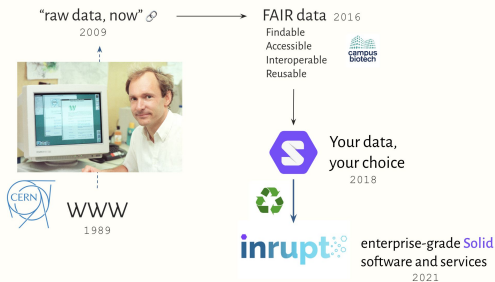
Improve privacy

+ Reduce data transaction costs

+ Increase data sharing

→ Open knowledge





enterprise-grade Solid
software and services

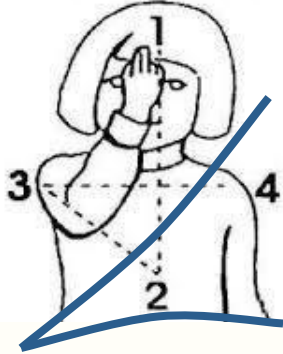


The healthy option for companies
to spend their budget.

SUSTAINABLE OPEN DATA

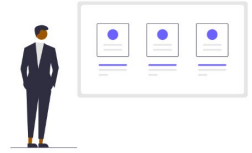
4i Ai4Privacy

4 EASY STEPS FOR GDPR COMPLIANCE



Company Mission

“Manage privacy by default:
simple, effective and human-centric.”



Social Impact

“Building the 21st century seatbelt for data privacy”

SOURCES: [Open source seatbelt patent](#)

AI-Privacy Protection with 3-Level Process: Demo

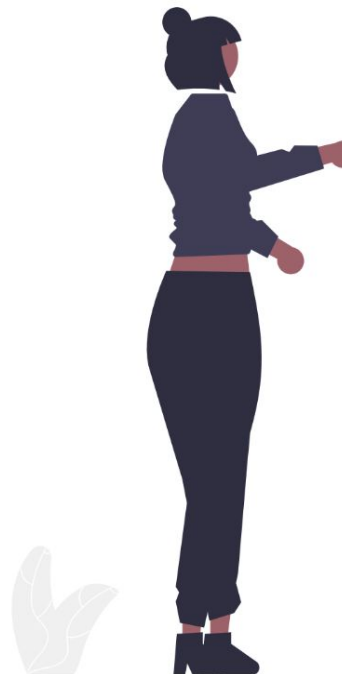
DEMO 

Ms. **PREFIX** Johnson **LASTNAME** at only
25 **AGE** is one of our top customers,
possessing several properties in **Regent**
Street **STREET**, **London** **STATE**, the majority
shares of **Turnip GmbH** **COMPANYNAME** and
a life insurance.

✓ 1. **Awareness**
Privacy report

✓ 2. **Protect**
Privacy mask

✓ 3. **Enable**
Privacy unmask



Dataset Viewer

Auto-converted to Parquet
API
View in Dataset Viewer

Split (2)
train · 178k rows

Search this dataset

source_text	target_text	privacy_mask	span_labels
string · lengths	string · lengths	list · lengths	string · lengths
 36512	 8848	 079	 2
Subject: Group Messaging for Admissions Process Good morning, everyone, I hope th...	Subject: Group Messaging for Admissions Process Good morning...	[{ "value": "wynqvrh053", "start": 287, "end": 297, "label": "USERNAME" }, { ...	[[440, 453, "USERNAME"]],
- Meeting at 2:33 PM - N23 - Meeting at 11:29pm - wennmann27 - Meeting at 4:45 PM...	- Meeting at [TIME] - [USERNAME] - Meeting at [TIME] - [USERNAME]	[{ "value": "2:33 PM", "start": 13, "end": 20, "label": "TIME" }, { "value": "...	[[74, 81, "T [50, 60, ...
Subject: Admission Notification - Great Britain University\nDear...	Subject: Admission Notification - Great Britain University\nDear...	[{ "value": "5:24am", "start": 263, "end": 269, "label": "TIME" }, { "value": "...	[[395, 407, "SOCIALNUMBE
Card: KB90324ER\n Country: GB\n Building: 163\n Street: Conygre Grove\n City:...	Card: [IDCARD]\n Country: [COUNTRY]\n Building:...	[{ "value": "KB90324ER", "start": 6, "end": 15, "label": "IDCARD" }, { "value": "...	[[390, 393, "STATE"], [3
N, WA14 5RW\n Password: r]iD1#8\n...and so on for all applicants...	N, WA14 5RW\n Password: [PASS]\n...and so on for all...	[{ "value": "r]iD1#8", "start": 26, "end": 33, "label": "PASS" }, { "value": "...	[[336, 352, "DATE"], [26
Subject: Admission Application Attachments Confirmation Dear Applicants, We hope thi...	Subject: Admission Application Attachments Confirmation Dear...	[{ "value": "301025226", "start": 311, "end": 320, "label": "PASSPORT" }]	[[311, 320, "PASSPORT"]]

Downloads last month 1,318

</> Use in Datasets library

Edit dataset card

⋮

Size of downloaded dataset files:

803 MB

Size of the auto-converted Parquet files:

197 MB

Number of rows:

225,405



Hugging Face

Dataset

AI4PRIVACY OPEN DATASETS

#open-source

LLM Status Codes

HTTP-like codes for LLMs

4



[LLM status codes](#) 



Michael Anthony

- Co-founder, CEO
- Masters of AI in Switzerland USI / ETH

ETH zürich



Silvia Pareti

- Co-founder, CPO
- 9.5+ years at Google in Natural Language Processing

Google



Valerio Battaglia

- Co-founder, CTO
- Ex-Google Partner Engineer in operations

Google



Konstantinos Samaras-Tsakiris

- Research and Technical advisor
- Ex-CERN Researchers



Chris Phillips

- Commercial advisor
- 25 years entrepreneur in translations

argos
multilingual

4ⁱ Ai4Privacy TEAM