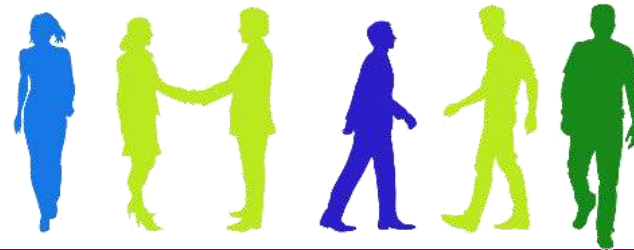


Journal Club - GESAN

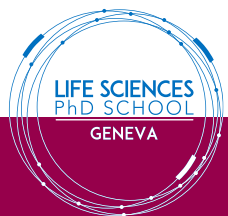


Some challenges and hints handling personal human data in data science

Christian Lovis
Division of medical information sciences



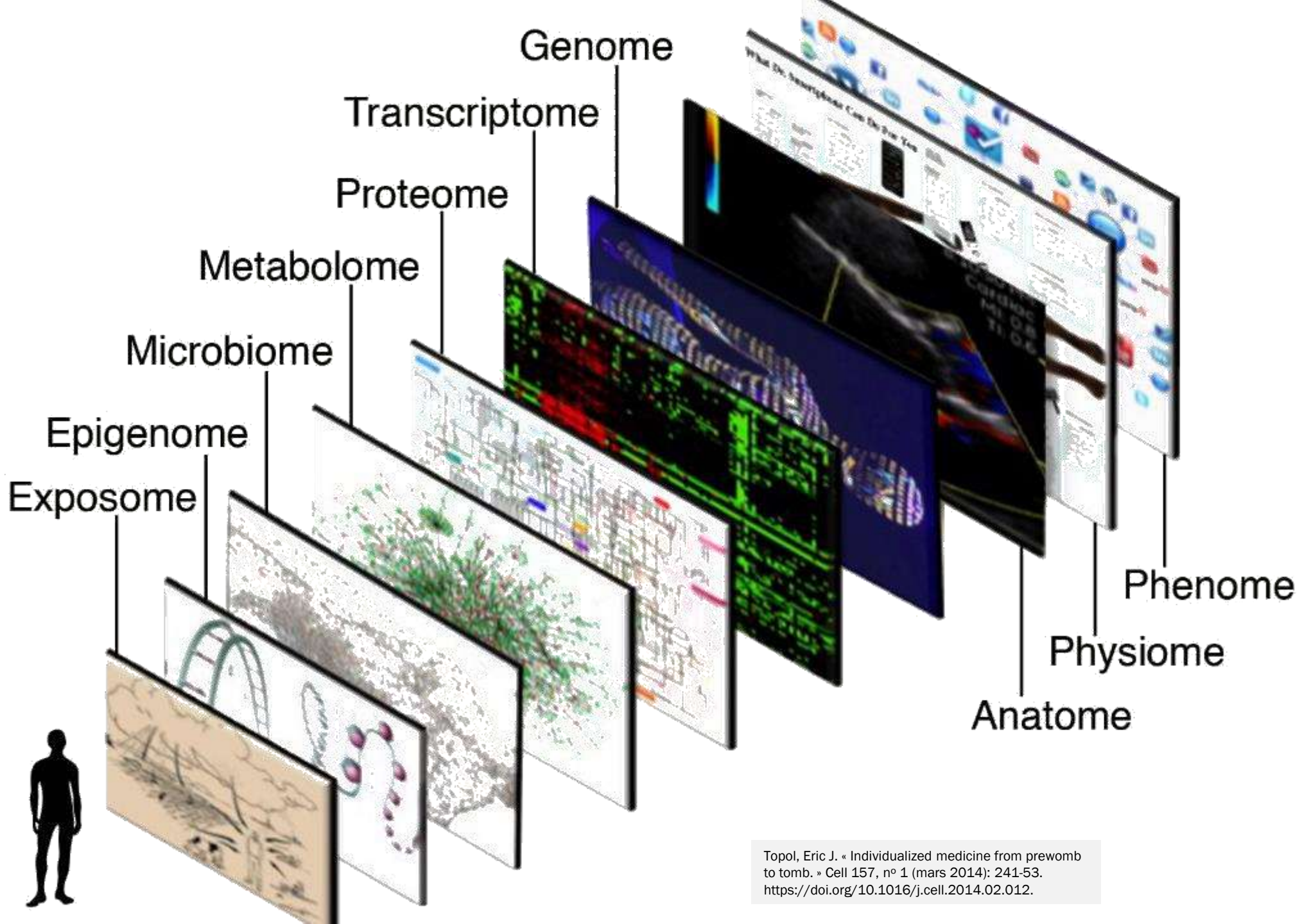
Hôpitaux
Universitaires
Genève



Genomics & Digital Health

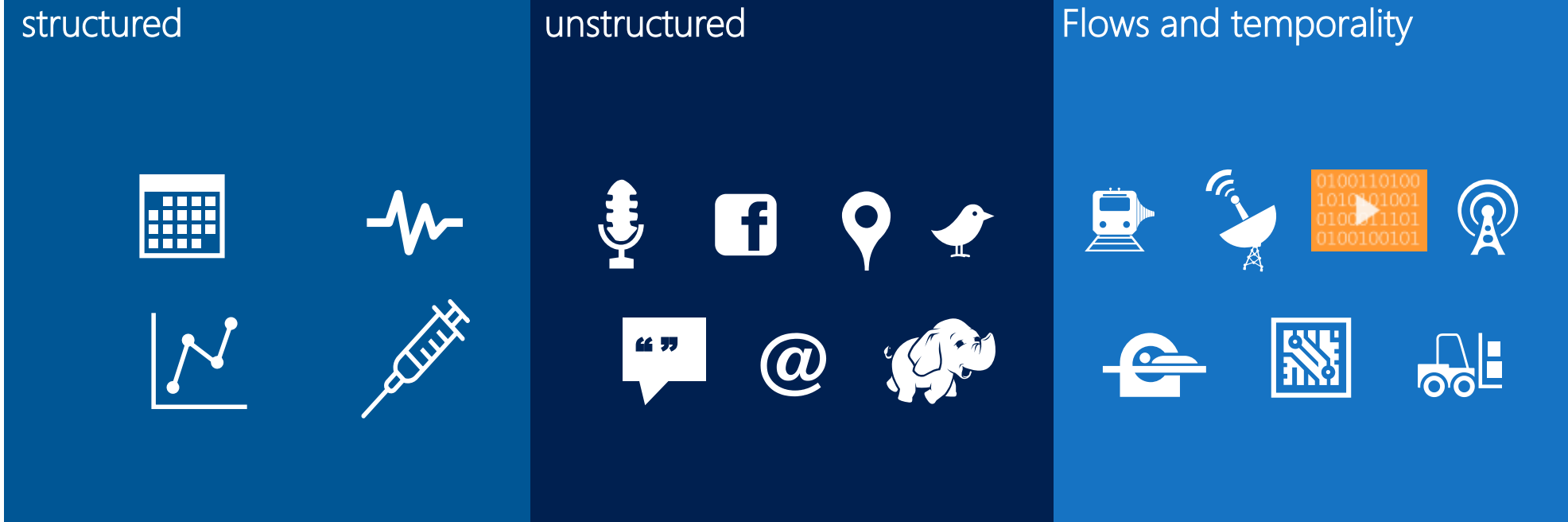


UNIVERSITÉ
DE GENÈVE



MULTIMODAL – HETEROGENOUS – DISTRIBUTED



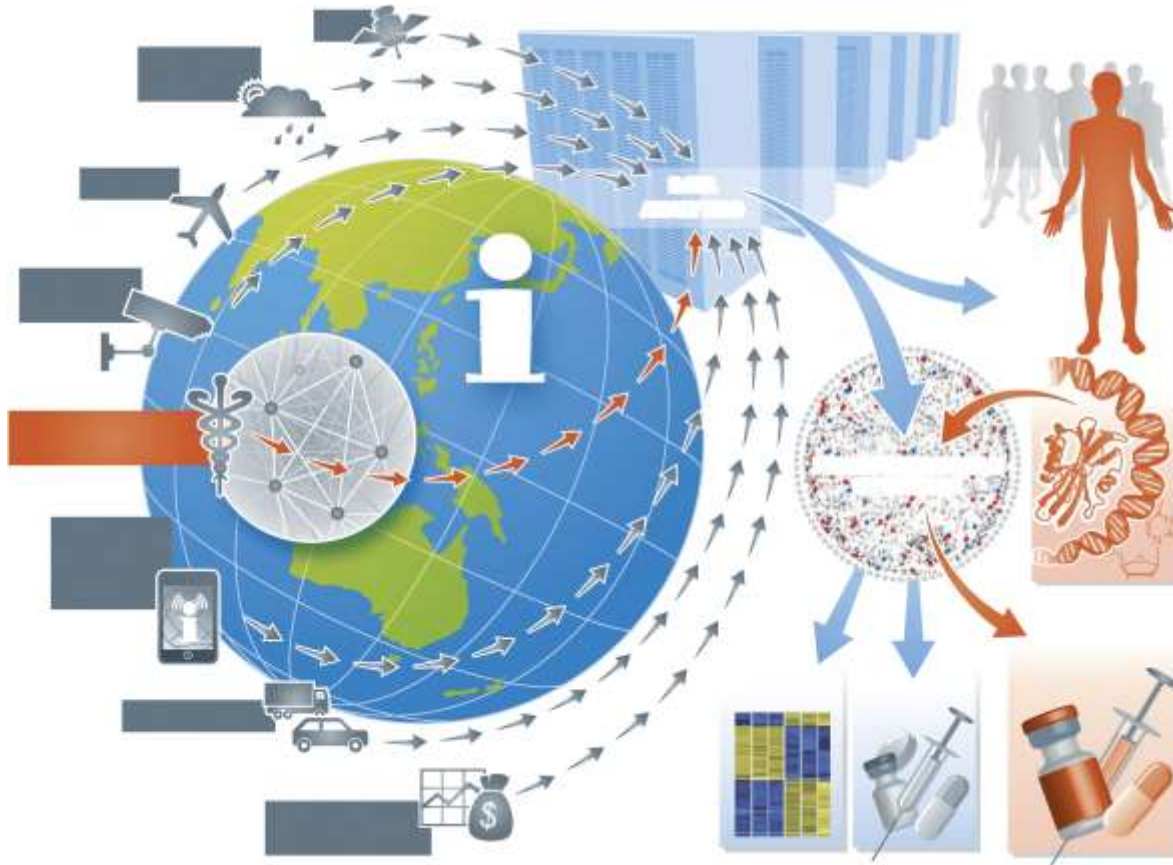


- ▶ all data around, complex data types, graph relations
- ▶ temporality
- ▶ Flows and streams

EDITORIAL

The changing privacy landscape in the era of big data

Molecular Systems Biology 8: 612; published online 11 September 2012; doi:10.1038/msb.2012.47



Big data are all around us, enabled by technological advances in micro- and nano-electronics, nano materials, interconnectivity provided by sophisticated telecommunication infrastructure, massive network-attached storage capabilities, and commodity-based high-performance computing infrastructures. The ability to store all credit card transactions, all cell phone traffic, all e-mail traffic, video and images from extensive networks of surveillance devices, satellite and ground sensing data informing on all aspects of the weather and overall climate, and now to generate and store massive data informing on our personal health including whole genome sequencing data and extensive imaging data, is driving a revolution in high-end data analytics to make sense of the big data, drive more accurate descriptive and predictive models that inform decision making on every level, whether identifying the next big security threat or making the best diagnosis and treatment choice for a given patient.

Does PatientsLikeMe sell my information?

Yes, we do. We create partnerships between you, our patients, and the companies that are developing products to help you. To do that, we take the information you entrust to us and sell it to the companies that can use that data to improve or understand products or the disease market. The personally identifiable information you share upon joining the site (like your name or contact information) is considered restricted and therefore not shared with partners. The data and text you enter in and around the shared parts of the site (e.g., on your profile, in the forum, symptom or treatment reports) may be shared or sold in aggregate to partners.

We also provide a voluntary opt-in service to allow partners to directly communicate with patient members through our system.

Related articles

- [How does PatientsLikeMe make money?](#)
- [What sort of information should I share in the About Me section?](#)

- La semaine 21 compte un jour férié (lundi de Pentecôte), qui rend les comparaisons difficiles.
- Si l'on considère uniquement les jours ouvrables de mardi à vendredi, les chiffres montrent une fréquentation équivalente à environ 74% de la fréquentation 2019.

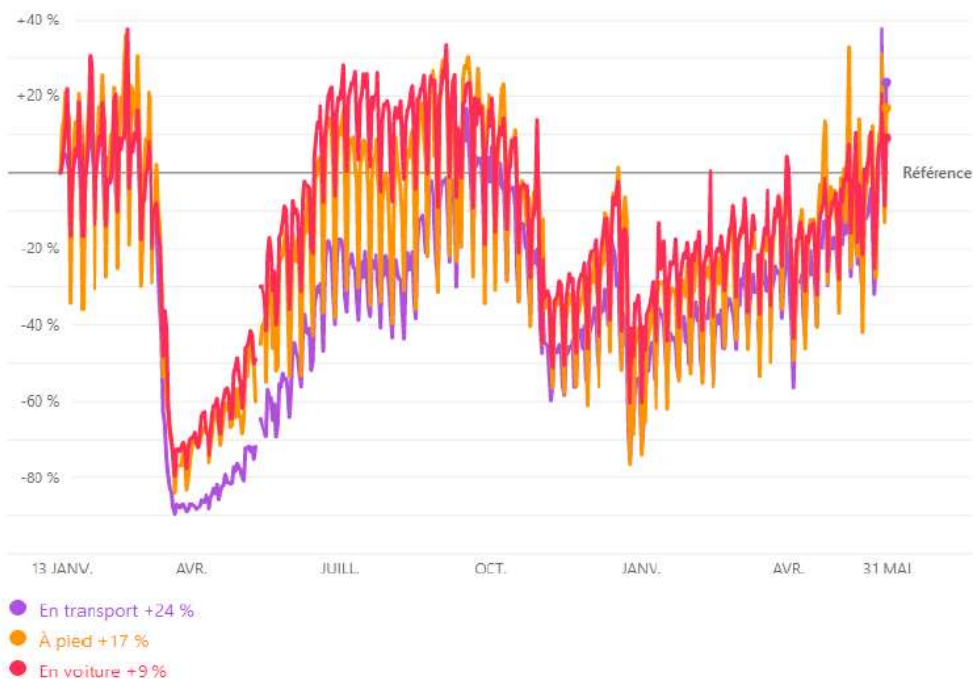


Figure 25 COVID-19, mobilité Apple – évolution des demandes d'itinéraires depuis le 13 janvier 2020 sur Apple Plans, à Genève

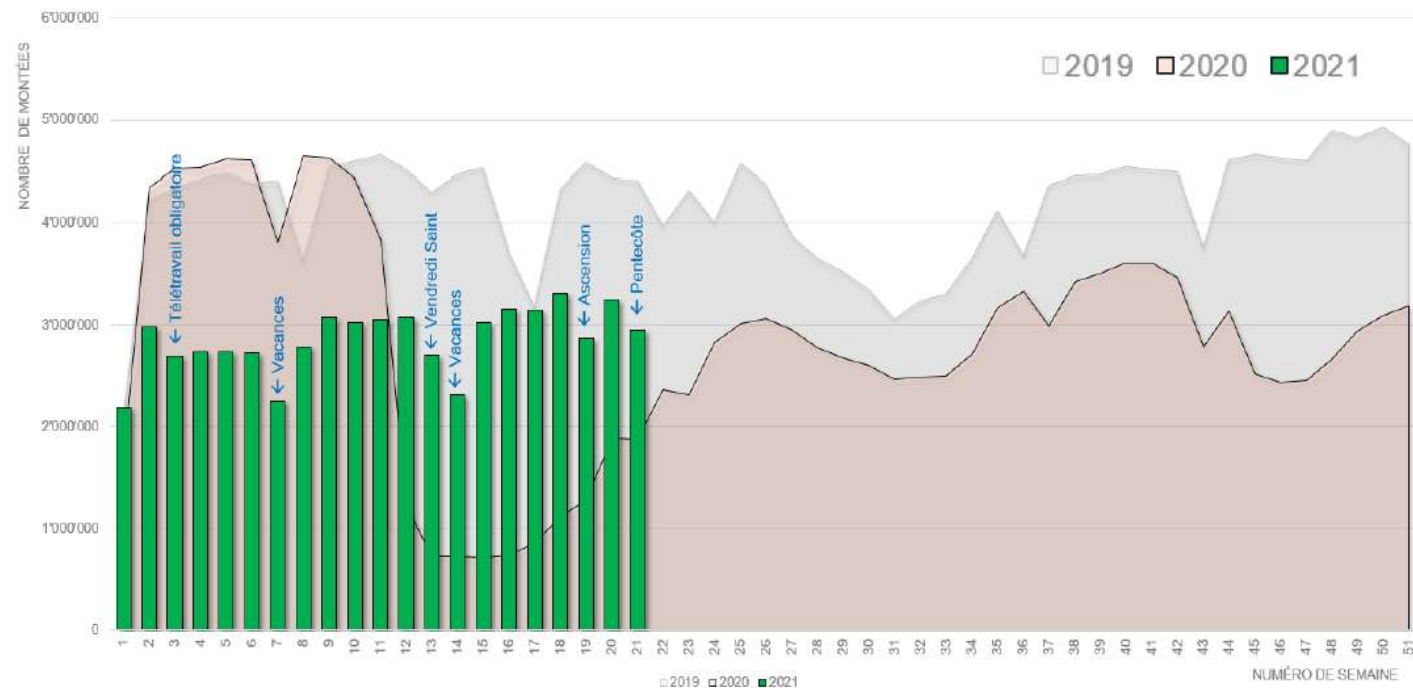


Figure 24 COVID-19, mobilité TPG – nombre de montées par semaine sur les lignes TPG principales, secondaires et transfrontalières

- Sept semaines après les vacances de Pâques, augmentation nette de la mobilité au dessus du seuil de référence.

Ces données proviennent notamment de sources externes (Apple) sans disponibilité des données brutes.

World

U.S. soldiers are revealing sensitive and dangerous information by jogging



1.0k



GPS tracking company Strava published an interactive map in Nov. 2017, showing where people have used fitness tracking devices.
(Patrick Marlin/The Washington Post)

By **Liz Sly** January 29 [Email the author](#)

BEIRUT — An interactive map posted on the Internet that shows the whereabouts of people who use fitness devices such as Fitbit also reveals highly sensitive information about the locations and activities of soldiers at U.S. military bases, in what appears to be a major security oversight.

The [Global Heat Map](#), published by the GPS tracking company Strava, uses satellite information to

Under Investigation

Archive

Help for Consumers

As required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals. The following breaches have been reported to the Secretary:

Cases Currently Under Investigation

This page lists all breaches reported within the last 24 months that are currently under investigation by the Office for Civil Rights.

[Show Advanced Options](#)

Breach Report Results							
Expand All	Name of Covered Entity	State	Covered Entity Type	Individuals Affected	Breach Submission Date	Type of Breach	Location of Breached Information
	Forefront Dermatology, S.C.	WI	Healthcare Provider	2413553	07/08/2021	Hacking/IT Incident	Network Server
	St. Joseph's/Candler Health System, Inc.	GA	Healthcare Provider	1400000	08/10/2021	Hacking/IT Incident	Network Server
	University Medical Center Southern Nevada	NV	Healthcare Provider	1300000	08/13/2021	Hacking/IT Incident	Network Server
	Professional Business Systems, Inc., d/b/a Practicefirst Medical Management Solutions and PBS Medcode Corp., ("Practicefirst")	NY	Business Associate	1210688	07/01/2021	Hacking/IT Incident	Network Server
	UF Health Central Florida	FL	Healthcare Provider	700981	07/30/2021	Hacking/IT Incident	Network Server
	DuPage Medical Group, Ltd.	IL	Healthcare Provider	655384	08/30/2021	Hacking/IT Incident	Network Server
	UNM Health	NM	Healthcare Provider	637252	08/03/2021	Hacking/IT Incident	Network Server
	State of Alaska Department of Health & Social Services	AK	Health Plan	500000	09/22/2021	Hacking/IT Incident	Desktop Computer, Laptop, Network Server
	Orlando Family Physicians, LLC	FL	Healthcare Provider	447426	07/20/2021	Hacking/IT Incident	Email
	Denton County, Texas	TX	Healthcare Provider	328417	08/24/2021	Unauthorized Access/Disclosure	Network Server
	USV Optical, Inc.	NJ	Healthcare Provider	180000	09/03/2021	Hacking/IT Incident	Network Server
	Metro Infectious Disease Consultants	IL	Healthcare Provider	171740	08/16/2021	Hacking/IT Incident	Email
	Simon Eye Management	DE	Healthcare Provider	144373	09/14/2021	Hacking/IT Incident	Email
	HealthReach Community Health Centers	ME	Healthcare Provider	122340	07/06/2021	Improper Disposal	Electronic Medical Record

11'893'247

01.07.2021 – 05.10.2021

https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

HEALTHCARE

Abundance of stolen healthcare records on dark web is causing a price collapse

**SHARE****WRITTEN BY**

Chris Bing

OCT 24, 2016 | CYBERSCOOP

The large amount of leaked patient records stolen and posted for sale to the dark web in recent months has caused prices for most of those records to drop, according to new research provided to CyberScoop and conducted by the Institute for Critical Infrastructure Technology and cybersecurity firms Flashpoint and Intel Security.

In the face of exceeding supply, stagnant demand and increased law enforcement attention, it's becoming increasingly difficult for criminals to make a living selling partial healthcare records, according to James Scott, a senior fellow at ICIT.

While the quality, quantity and sometimes origin of such electronic records will help dictate the price of any specific package for sale, average prices are largely trending downwards for individual, non-financial files, new research shows. The value of similar healthcare records that sold last year for roughly \$75 to \$100 dollars can now be found for around \$20 to \$50 dollars, Scott **said**.



Health & Science

FDA, facing cybersecurity threats, tightens medical-device standards



An artificial cadaver that is used by the Security and Privacy Research Lab at the University of Michigan. Researchers use the artificial cadaver to test the security and privacy of various medical devices, including pacemakers and defibrillators. (Joseph Xu/Michigan Engineering Communications & Marketing)

By **Lena H. Sun** and **Brady Dennis** June 13, 2013 Email the author

The Switch

A new hacker money-making strategy: Betting against insecure companies on Wall Street

By **Andrea Peterson** September 1, 2016

For decades, there's been an unofficial truce between cybersecurity researchers and companies: When good guy hackers find a problem, they give companies a chance to fix it before going public.

But a cybersecurity firm called MedSec just upended that truce.

Instead of following industry traditions by alerting St. Jude Medical when researchers found alleged bugs in the company's implantable heart equipment, MedSec struck a deal with a short-seller called Muddy Waters Research. The investment firm would make the vulnerabilities public in exchange for giving the cybersecurity firm a cut of the profits Muddy Waters made from betting against the medical device maker's stock, MedSec chief executive Justine Bone said in an interview. The arrangement was first reported by [Bloomberg News](#).



A Major Drug Company Now Has Access to 23andMe's Genetic Data. Should You Be Concerned?

By [Jamie Ducharme](#)

July 26, 2018

TIME Health



Photo courtesy of 23andMe

Consumer genetic testing company 23andMe [announced on Wednesday](#) that GlaxoSmithKline purchased a \$300 million stake in the company, allowing the pharmaceutical giant to use 23andMe's trove of genetic data to develop new drugs — and raising new privacy concerns for consumers.

The “collaboration” is a way to make “novel treatments and cures a reality,” [23andMe CEO Anne Wojcicki](#) said in a company blog post. But, though [it isn't 23andMe's first foray into drug discovery](#), the deal doesn't seem quite so simple to some medical experts — or some of the roughly 5 million 23andMe customers who have sent off tubes of their spit in exchange for ancestry and health insights.

Perhaps the most obvious issue is privacy, says Peter Pitts, president of the Center for Medicine in the Public Interest, a non-partisan non-profit that aims to promote patient-centered health care.

Sloan Kettering's Cozy Deal With Start-Up Ignites a New Upheaval



At Memorial Sloan Kettering Cancer Center in Manhattan, doctors and staff objected to a for-profit venture that could be lucrative for a few leading researchers and board members. Gabriella Angotti-Jones/The New York Times

By Charles Ornstein and Katie Thomas

Sept. 20, 2018



An artificial intelligence start-up founded by three insiders at Memorial Sloan Kettering Cancer Center debuted with great fanfare in February, with **\$25 million in venture capital** and the promise that it might one day transform how cancer is diagnosed.

The company, [Paige.AI](#), is one in a burgeoning field of start-ups that are applying artificial intelligence to health care, yet it has an advantage over many competitors: The company [has an exclusive deal to use the cancer center's vast archive](#) of 25 million patient tissue slides, along with decades of work by its world-renowned pathologists.

Memorial Sloan Kettering holds an **equity stake in Paige.AI**, as does a member of the cancer center's executive board, the chairman of its pathology department and the head of one of its research laboratories. Three other board members are investors.

The arrangement has sparked considerable turmoil among doctors and scientists at Memorial Sloan Kettering, which has intensified [in the wake of an investigation by ProPublica and The New York Times](#) into the failures of its chief medical officer, **Dr. José Baselga**, to disclose some of his financial ties to the health and drug industries in dozens of research articles. He [resigned last week](#), and Memorial Sloan Kettering's chief executive, Dr. Craig B. Thompson, announced a new task force on Monday to review the center's conflict-of-interest policies.



"The principles of data protection should therefore not apply to [...] personal data rendered **anonymous** in such a manner that the data subject is not or no longer identifiable." (Recital 26)

"**pseudonymisation**' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures [...]" (Definition 5)

Directive 95/46/EC (General Data Protection Regulation)

GDPR, the EU's new data protection law

Anonymous information: information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. **It is an irreversible state.**

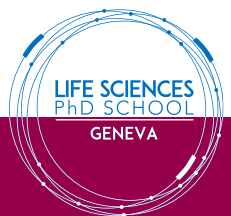
HHS.gov

U.S. Department of Health & Human Services

Health Information Privacy

Health Insurance Portability and Accountability Act of 1996 (HIPAA)

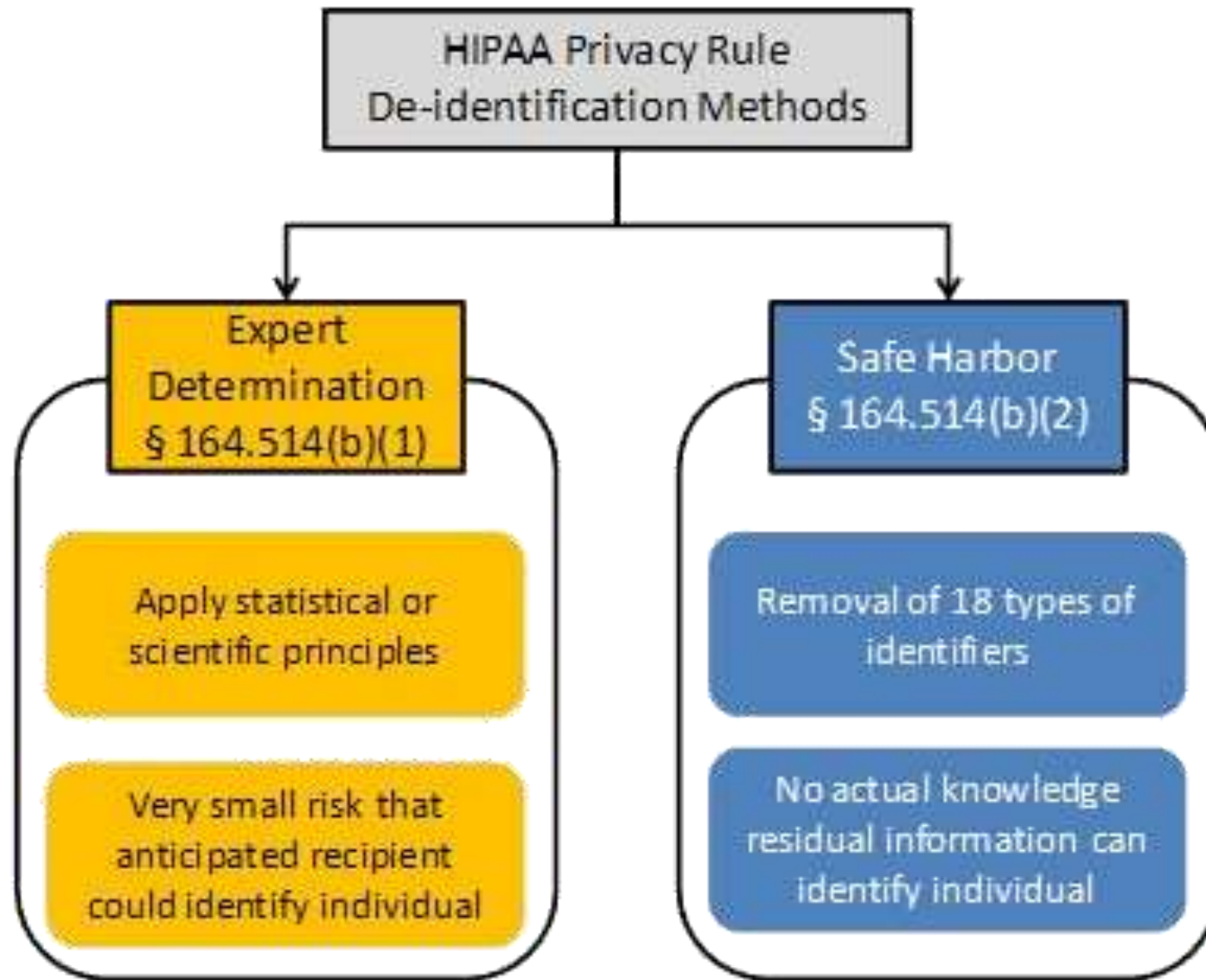
De-identification of PHI: health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual. Rule-based.



Genomics & Digital Health



UNIVERSITÉ
DE GENÈVE





Health Information Privacy

- **Names**
- All **geographic** subdivisions smaller than a state (street address, city, county, precinct, ZIP code, and their equivalent geocodes);
- All elements of **dates** (except year) directly related to an individual (birth date, admission date, discharge date, death date), and all ages over 89 and all elements of dates (including year) indicative of such age;
- **Telephone & Fax numbers;**
- **Device** identifiers and serial numbers;
- **Email** addresses;
- **Social security numbers;**
- **IP** addresses;
- **Medical record numbers;**
- **Biometric** identifiers, including finger and voice prints;
- **Health plan** beneficiary numbers.



10/01/2018

[Sloan Kettering Controversies: Trust is the Public Foundation of Medical Research](#)

by Ann Mongoven, PhD, MPH

1) **Research on human beings ethically requires their consent.**

This has been a core bioethical principle since research harms dramatized by the Nuremburg Trials and the Tuskegee Syphilis Experiment sparked regulations and processes to protect human research subjects. [...]

2) **Having conflicts of interest is not necessarily bad. Hiding them is.**

The Sloan missteps underscore the ethical importance of routine conflict-of-interest disclosure: to funders, to IRBs, to publishers, to boards, and to patients recruited to participate in research.

3) **The relationship between public good, private benefit, and non-profit status is complex.**

4) **Public trust is a prerequisite resource for health research.**

The twin Sloan cases remind us that the most fundamental resource necessary for any large-scale health research is public trust.

This entry was posted in [Clinical Trials & Studies](#), [Conflict of Interest](#), [Featured Posts](#), [Health Care](#), [Human Subjects Research & IRBs](#), [Informed Consent](#) and tagged [IRB](#), [Public Trust](#), [Sloan Kettering](#). Posted by [Ann Mongoven](#).

<http://www.bioethics.net/2018/10/sloan-kettering-controversies-trust-is-the-public-foundation-of-medical-research/>

pseudonimized
deidentified
anonymized



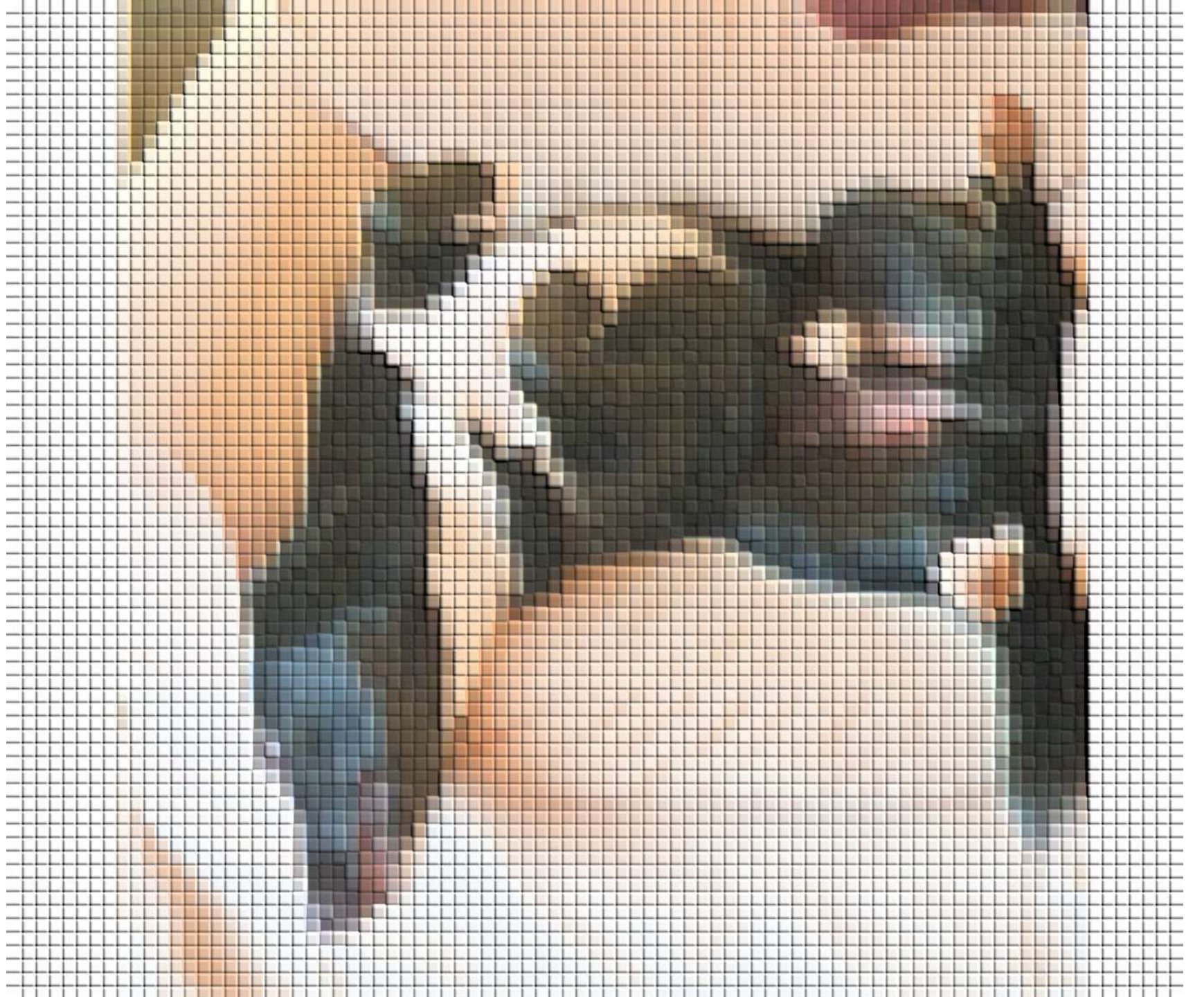
Babou



Babou



Babou



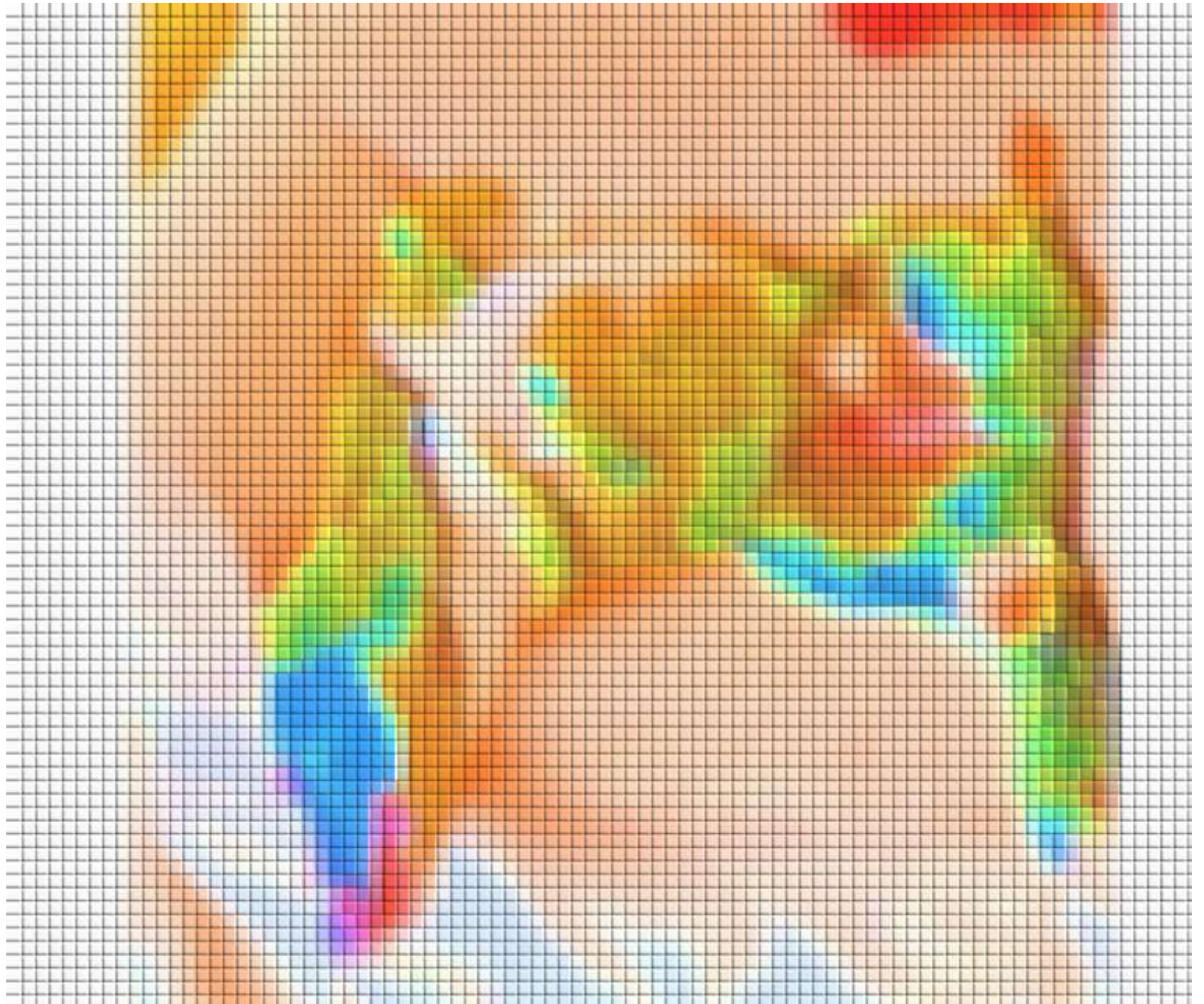
Babou



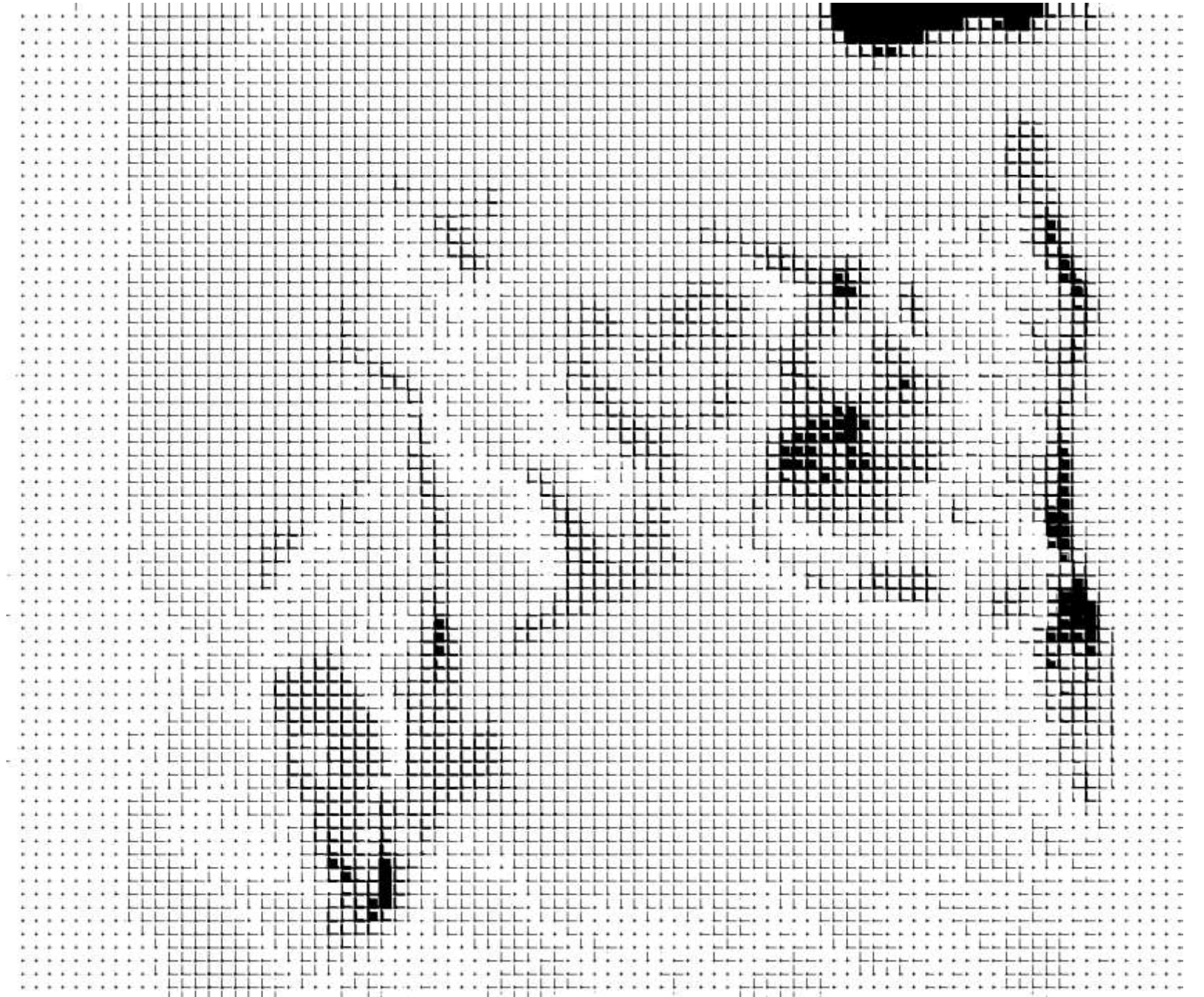
Babou



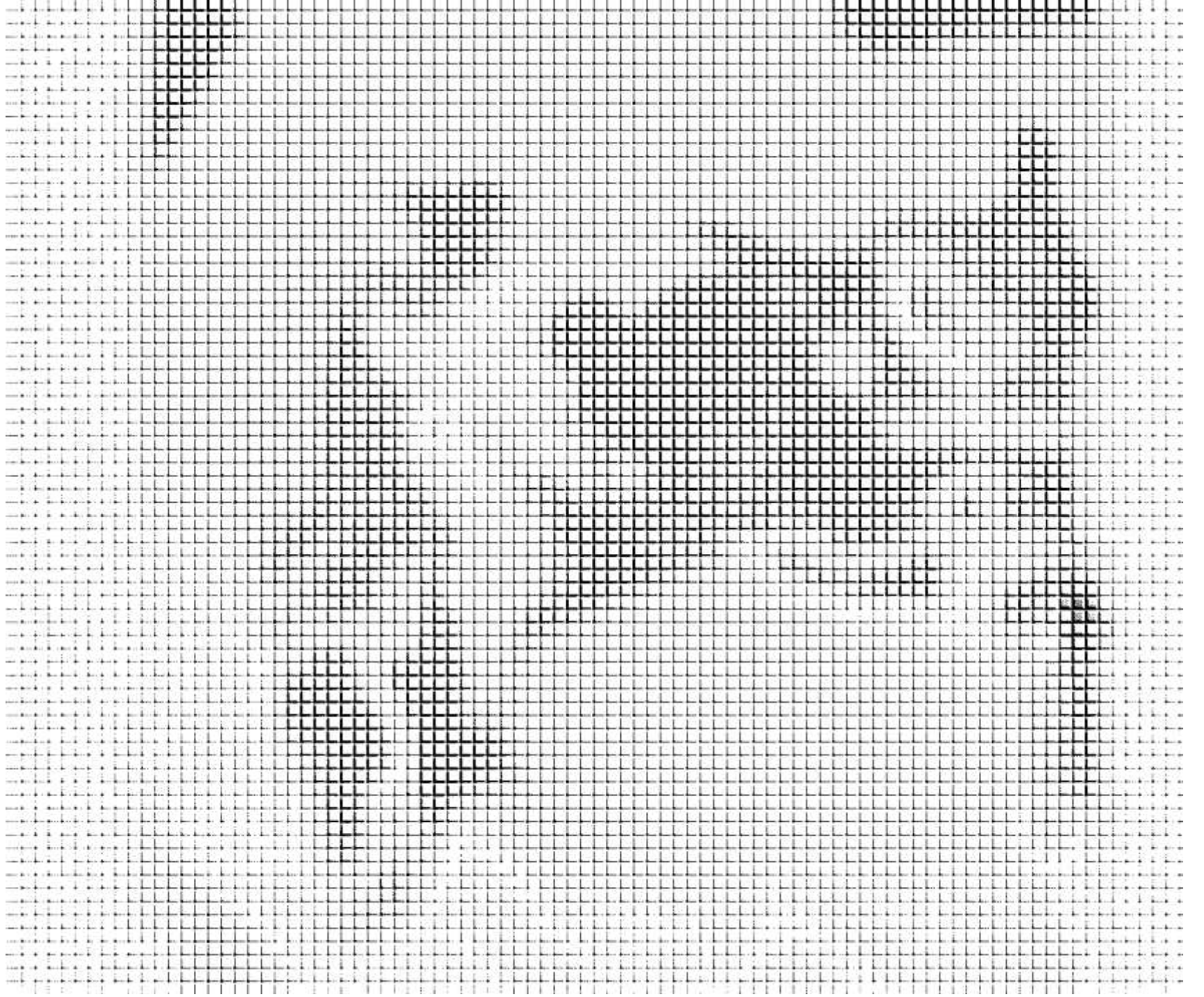
Babou



Babou



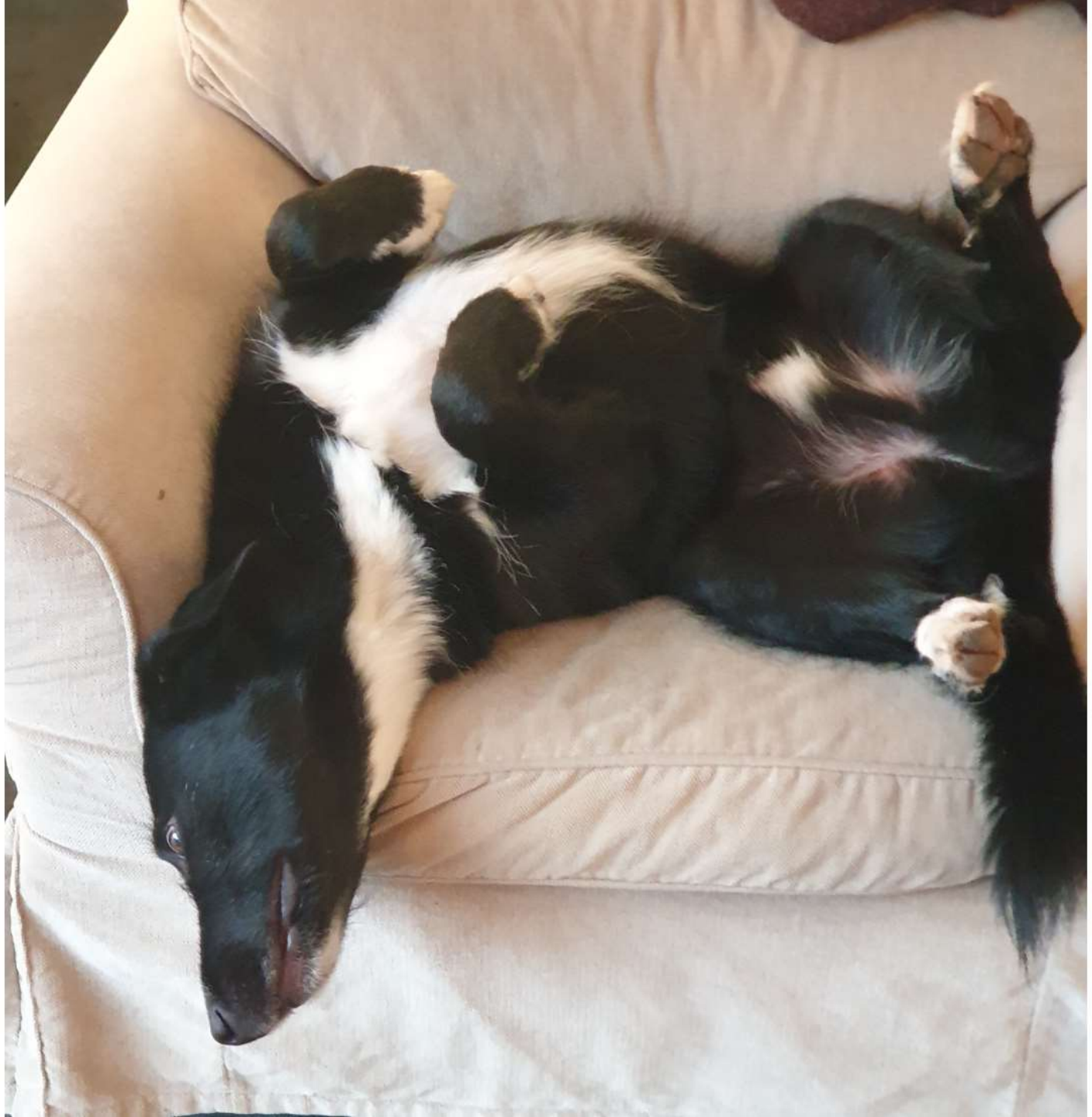
Babou



Babou



Jules



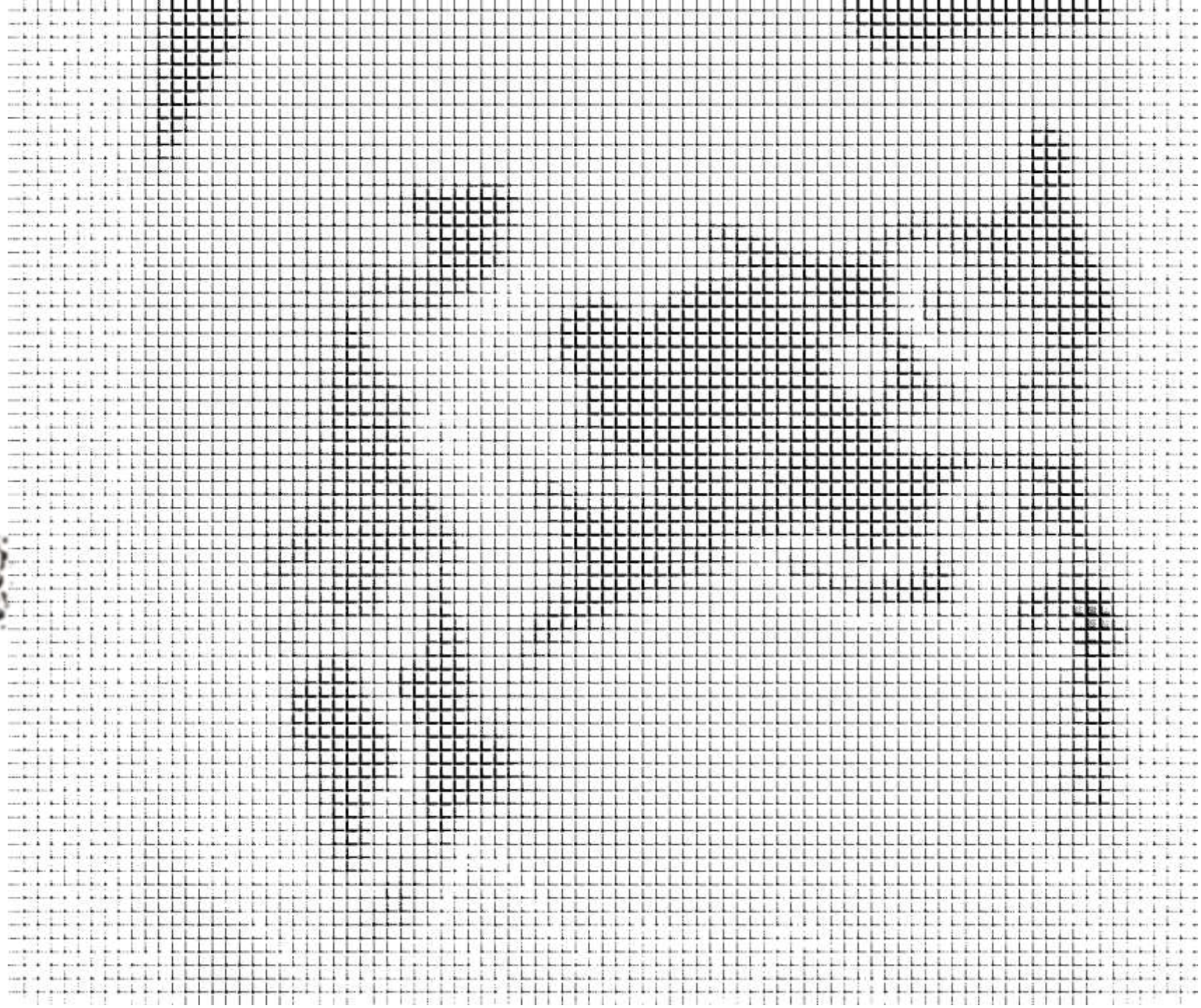
A\$Tfs_



Babou







Babou



KB approaches

Rule-based approach

- explainable
- correctable
- accountable

ML based approach

No existing anonymization algorithm provides both perfect privacy protection and perfect analytic utility. A spectral basis derived from the data's eigenvectors is one that can provide substantial improvement.

6.3 Open Problems

Anonymizing in kernel space

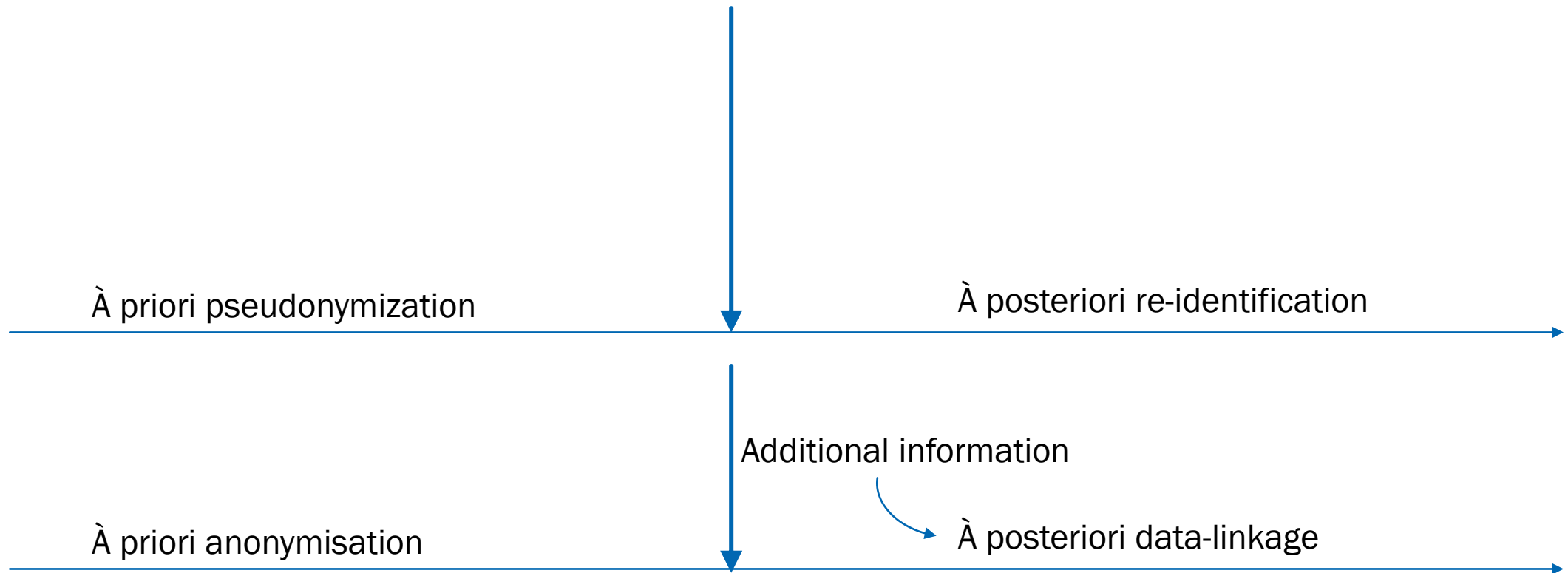
unacceptably distort the information we'd like to preserve in anonymization

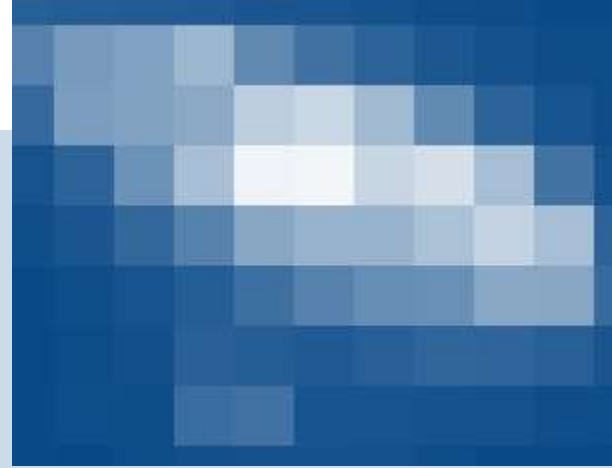
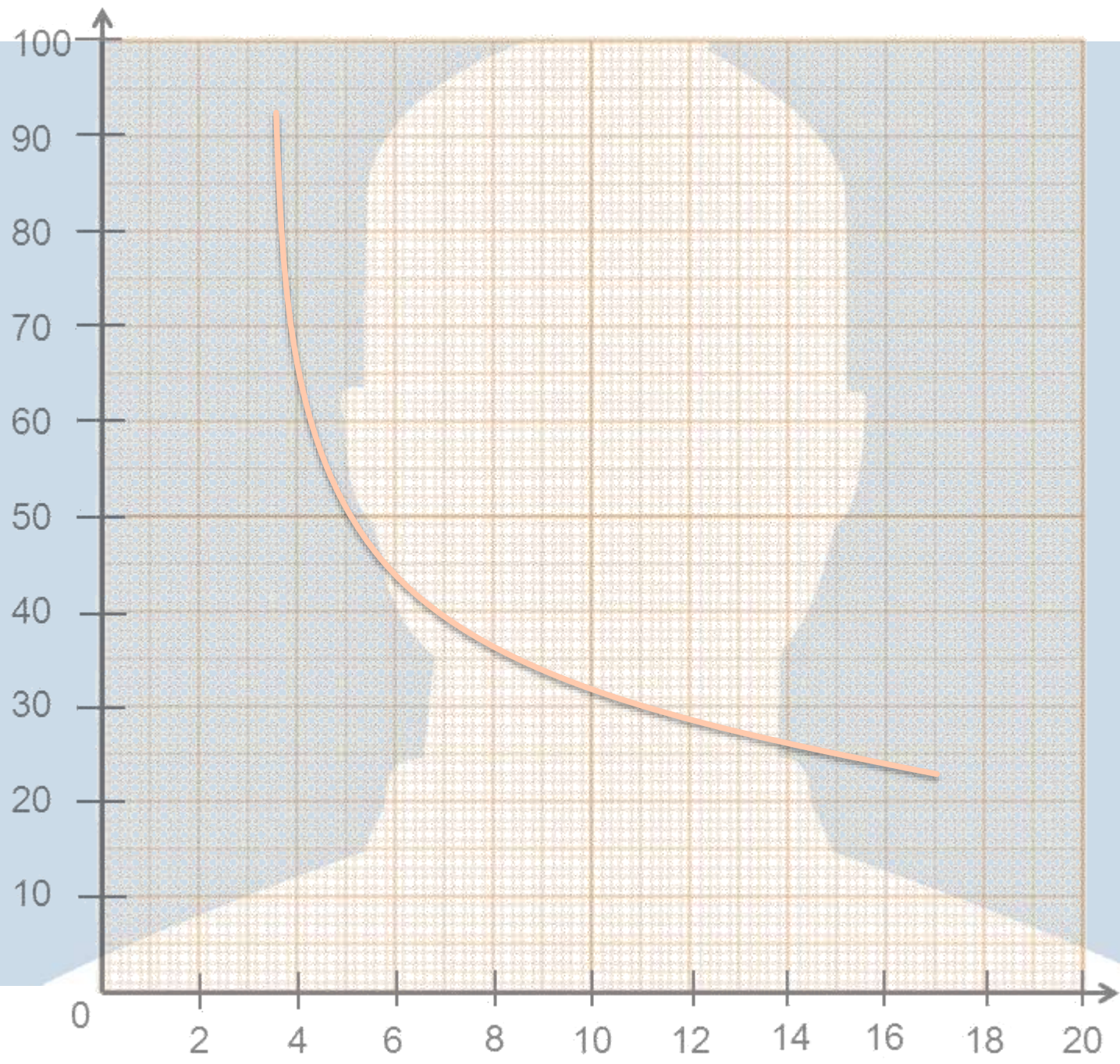
Time-series data

Hierarchical data

Very high dimensional data

ALL IS DATA





Reducing patient re-identification risk for laboratory results within research datasets

Ravi V Atreya,¹ Joshua C Smith,¹ Allison B McCoy,² Bradley Malin,^{1,3}
 Randolph A Miller^{1,4,5}

J Am Med Inform Assoc 2013;**20**:95–101. doi:10.1136/amiajnl-2012-001026

Table 1 Uniqueness for four, five, and six consecutive results of 10 representative laboratory tests and two panels (CBC and CHEM7)

Individual laboratory tests			
Test name	Number of consecutive laboratory tests and proportion unique (with number of items qualifying for analysis)		
	4	5	6
pH	0.590 (N=83 941)	0.937 (N=75 954)	0.994 (N=69 175)
Gluc	0.996 (N=133 259)	1.000 (N=110 669)	1.000 (N=93 693)
Ca	0.723 (N=51 905)	0.974 (N=41 140)	0.998 (N=33 283)
LymAbs	0.986 (N=27 591)	0.998 (N=22 361)	1.000 (N=18 615)
PCV	0.195 (N=201 941)	0.575 (N=172 619)	0.886 (N=149 514)
PT-INR	0.343 (N=47 768)	0.559 (N=38 875)	0.725 (N=32 188)
Chol	1.000 (N=780)	1.000 (N=613)	1.000 (N=478)
SGPT	0.996 (N=12 655)	0.999 (N=9807)	1.000 (N=7850)
CK	0.963 (N=6509)	0.979 (N=3659)	0.986 (N=2219)
Alb	0.649 (N=11 520)	0.924 (N=8606)	0.989 (N=6580)
Panel name	Laboratory panels		
CBC (five components)	0.988 (N=211 777)		
CHEM7 (seven components)	0.989 (N=239 253)		

Alb, albumin; Ca, calcium; CBC, complete blood count; CHEM7, blood test measuring electrolytes, glucose, and renal function; Chol, cholesterol; CK, creatine kinase; Gluc, glucose; LymAbs, absolute lymphocytes; PCV, packed cell volume; PT-INR, international normalized ratio for prothrombin time; SGPT, serum glutamic pyruvic transaminase.



A Systematic Review of Re-Identification Attacks on Health Data

Khaled El Emam^{1,2*}, Elizabeth Jonker¹, Luk Arbuckle¹, Bradley Malin^{3,4}

December 2011 | Volume 6 | Issue 12 | e28071

1 Electronic Health Information Laboratory, CHEO Research Institute, Ottawa, Canada, **2** Department of Paediatrics, University of Ottawa, Ottawa, Canada, **3** Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, United States of America, **4** Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee, United States of America

[Elife](#). 2017 Nov 28;6. pii: e27798. doi: 10.7554/eLife.27798.

Rapid re-identification of human samples using portable DNA sequencing.

[Zaaijer S](#)^{1,2}, [Gordon A](#)², [Speyer D](#)^{1,2}, [Piccone R](#)³, [Groen SC](#)⁴, [Erlich Y](#)^{1,2,5}.

[J Bioeth Inq](#). 2017 Sep 14. doi: 10.1007/s11673-017-9806-9. [Epub ahead of print]

Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?

[Phillips M](#)¹, [Dove ES](#)², [Knoppers BM](#)³.