
VLDB'19

*Los Angeles, August 2019*

Nicolas Anciaux, Luc Bouganim,
Philippe Pucheral, Iulian Sandu Popa
and Guillaume Scerri

**PETRUS team, Inria & UVSQ**
https://team.inria.fr/petrus/

# Personal Database Security and Trusted Execution Environments: A Tutorial at the Crossroads

# Personal data…
# … at the crossroads of Business and Privacy

## From the business perspective…

Personalized services (e.g., personalized searches, pay-as-you-xxx),

… and needed optimizations (e.g., energy consumption, network …),

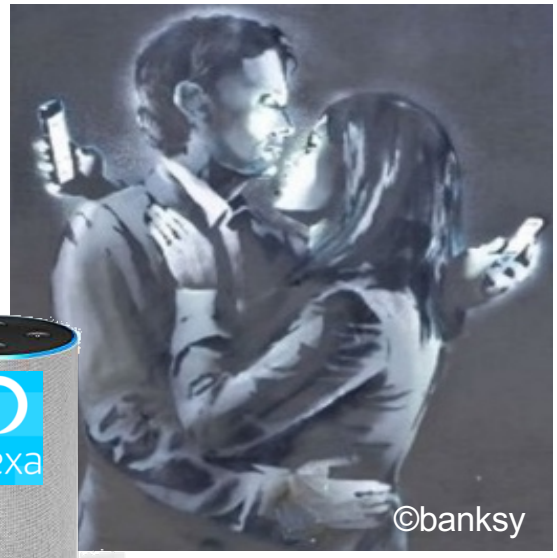
Various features improving business

… like targeted ads, improved CRM, increased time spend in social medias and games, etc.

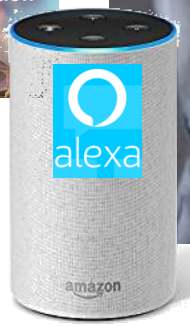
Ultimate profiling


*Source: crackedlabs.org*

# Personal data…
# … at the crossroads of Business and Privacy








## to societal concerns…

**Silent over-collection of personal data**

Eg: corp. (Alexa, Fortnite),gov. (Health Data Hub)

**Recurrent/massive leaks & attacks**

Eg: Yahoo, Equifax, Cambridge Analytica…

**Anonymised datasets often not anonymous**

Eg.: 15 fields is enough [RHM19]

**Uses considered questionable**

Eg: Social medias (Visa, Insurance)

Personal reports (Pipl, Intelius…),

…

**Discriminatory uses of personal data**

Eg: criterias in targeted ads,

e-justice, recruiting process

23andMe vs. GINA, …

THE WALL STREET JOURNAL.
New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed

pipl FILL IN THE BLANKS IN YOUR CUSTOMER LIST

This ad is for white people only.*

# Personal data…
# … at the crossroads of Business and Privacy

## … more advocacy of privacy issues & more acceptance by economic actors

### Legislation

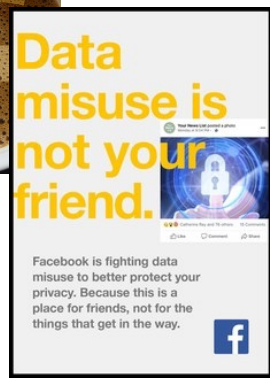
GDPR, Facial recognition forbidden in SF,
California Consumer Privacy Act (CCPA),
With fines applied




### More acceptance

Symptoms of a crisis of consciousness (e.g., Time well spent)
From "**privacy is no longer the social norms**"
… to "**private is the future**"
Privacy-based marketing campaigns

*Pop-up Cafés (?)*




*Chicago*


*New York*


*Toronto*


*Los Angeles*

# Current trend: give their personal data (agency) back to individuals

## Act I: the right to Data portability

… the right to retrieve its own data

## Act II: Personal Data Mg$^{t}$ Systems (PDMS)

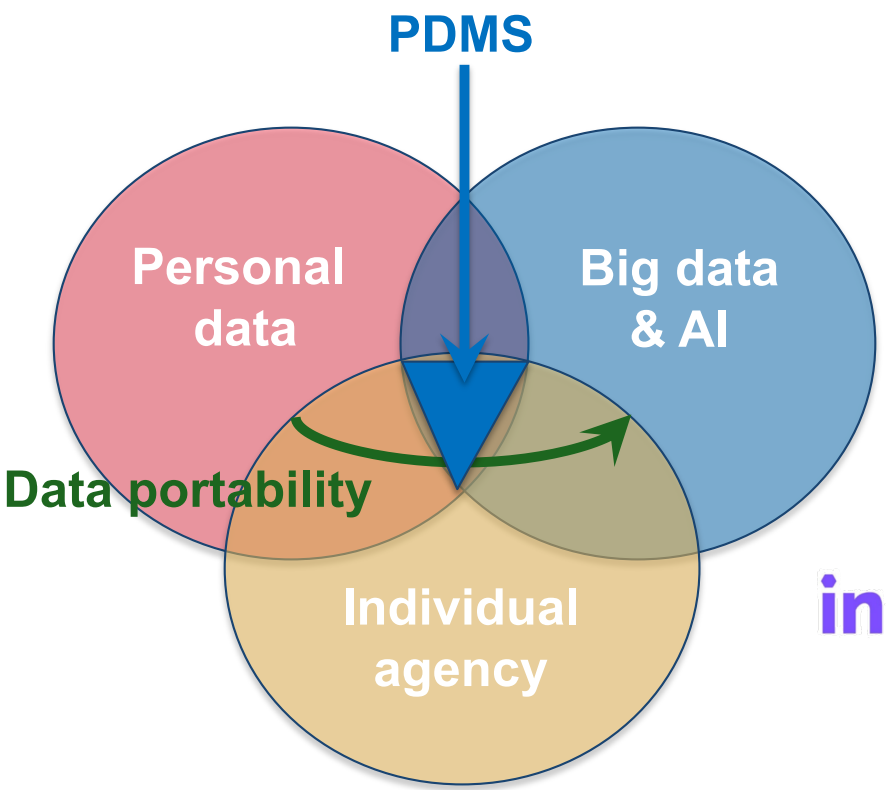
… the tool to manage its own data

Blue Button Download My Data

Green Button Connect My Data

MESINFOS

1 year GDPR




cozy.io

digi.me

DATABOX

BITSABOUT.ME

inrupt Tim Berners-Lee

# Is this enough to change the situation? …

TWO prerequisites !

**Individual's agency**

Let individuals freely decide about the new usages of their data all along their life cycle

Rather than: services in exchange of personal data

**Secured decentralized architectures**

Offer individuals the ability to securely control the raw data produced on their side

Rather than: centralizing everything in a few hands

# Is this enough to change the situation? …

**Individual's agency**

Let individuals freely decide about the new usages of their data all along their life cycle

Rather than: services in exchange of personal data

Major steps of personal data life-cycle escape today individual's control

Architectural considerations of a the PDMS platform are paramount

**Secured decentralized architectures**

Offer individuals the ability to securely control the raw data produced on their side

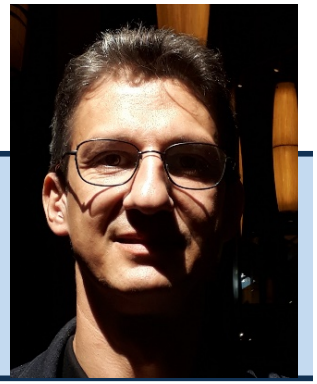
Rather than: centralizing everything in a few hands

Layman citizen …. as security expert?

Emergence of Trusted Execut$^{o}$ Env$^{t}$ (high-end servers & edges)

**The primary topic of this tutorial!**

# Tutorial Outline



## PART I. Personal Data Management Systems (PDMS)

Review of functionalities & addressed privacy threats

Individual's PDMS vs (corporate) DBMS and main properties to achieve

## PART II. TEE-based Data Management

The promises of Trusted Execution Environments (TEEs)

A review of privacy-preserving data management using TEEs

## PART III. Bridging the Gap between PDMS and TEEs

How could the main properties be achieved?

A quick view of remaining challenges

# References (1)


[AAB+10] T. Allard, N. Anciaux, L. Bouganim, Y. Guo, L. L. Folgoc, B. Nguyen, P. Pucheral, I. Ray, S. Yin. Secure personal data servers: a vision paper. PVLDB, 3(1), 25-35, 2010.

[ABB+19] N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, P. Pucheral, I. S. Popa, G. Scerri. Personal data management systems: The security and functionality standpoint. Inf. Syst., 80:13–35, 2019.

[ABD+19] M. Acosta, T. Berners-Lee, A. Dimou, J. Domingue, L-D. Ibá, K. Janowicz, M-E. Vidal, A. Zaveri: The FAIR TRADE Framework for Assessing Decentralised Data Solutions. WWW 2019

[ABP+14] N. Anciaux, L. Bouganim, P. Pucheral, Y. Guo, L. L. Folgoc, S. Yin. Milo-DB: a personal, secure and portable database machine. Distributed and Parallel Databases, 32(1):37–63, 2014.

[AEJ+15] A. Arasu, K. Eguro, M. Joglekar, R. Kaushik, D. Kossmann, R. Ramamurthy: Transaction processing on confidential data using cipherbase. ICDE 2015: 435-446

[AEK+14] A. Arasu, K. Eguro, R. Kaushik, R. Ramamurthy: Querying encrypted data. SIGMOD Conference 2014: 1259-1261

[AK13] A. Arasu, R. Kaushik: Oblivious Query Processing. ICDT 2014.

[ALS+15] N. Anciaux, S. Lallali, I. Sandu Popa, P. Pucheral: A Scalable Search Engine for Mass Storage Smart Objects. PVLDB 8(9): 910-921 (2015)

[ANS13] N. Anciaux, B. Nguyen, I. Sandu Popa: Personal Data Management with Secure Hardware: How to Keep Your Data at Hand. MDM (2) 2013: 1-2

[ANS14] N. Anciaux, B. Nguyen, I. Sandu Popa: Tutorial: Managing Personal Data with Strong Privacy Guarantees. EDBT 2014: 672-673

# References (2)


[BBB+17] R. Bahmani, M. Barbosa, F. Brasser, B. Portela, A.-R. Sadeghi, G. Scerri, B. Warinschi: Secure Multiparty Computation from SGX. Financial Cryptography 2017: 477-497

[BEE+17] J. Bater, G. Elliott, C. Eggen, S. Goel, A. Kho, J. Rogers: SMCQL: secure querying for federated databases. PVLDB 2017

[BGC+18] V. Bindschaedler, P. Grubbs, D. Cash, T. Ristenpart, V. Shmatikov: The tao of inference in privacy-protected databases. PVLDB 2018

[BPS+16] M. Barbosa, B. Portela, G. Scerri, B. Warinschi: Foundations of Hardware-Based Attested Computation and Application to SGX. EuroS&P 2016: 245-260

[BS11] S. Bajaj, R. Sion: TrustedDB: a trusted hardware-based database with privacy and data confidentiality. SIGMOD Conference 2011: 205-216

[DSC+15] T. T. A. Dinh, P. Saxena, E. Chang, B. C. Ooi, C. Zhang: M2R: Enabling Stronger Privacy in MapReduce Computation. USENIX Security 2015

[EZ17] S. Eskandarian, M. Zaharia: An oblivious general-purpose SQL database for the cloud. CoRR, abs/1710.00458, 2017

[FBB+18] B. Fuhry, R. Bahmani, F. Brasser, F. Hahn, F. Kerschbaum, A.-R. Sadeghi: HardIDX: Practical and secure index with SGX in a malicious environment. Journal of Computer Security 26(5): 677-706 (2018)

[HZX18] T. Hunt, Z. Zhu, Y. Xu, S. Peter, E. Witchel: Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data. ACM Trans. Comput. Syst. 35(4): 13:1-13:32 (2018)

# References (3)


[LAP+19] R. Ladjel, N. Anciaux, P. Pucheral, G. Scerri. Trustworthy Distributed Computations on Personal Data Using Trusted Execution Environments. TrustCom, 2019.

[LAS+17] S. Lallali, N. Anciaux, I. Sandu Popa, P. Pucheral: Supporting secure keyword search in the personal cloud. Inf. Syst. 72: 1-26 (2017)

[LSB19a] J. Loudet, I. Sandu Popa, L. Bouganim: SEP2P: Secure and Efficient P2P Personal Data Processing. EDBT 2019.

[LSB19b] J. Loudet, I. Sandu-Popa, L. Bouganim. DISPERS: Securing Highly Distributed Queries on Personal Data Management Systems. PVLDB 2019

[LWG+13] S. Lee, E.L. Wong, D. Goel, M. Dahlin, V. Shmatikov, πbox: A platform for privacy-preserving apps, in: NSDI, 2013.

[MPC+18] P. Mishra, R. Poddar, J. Chen, A. Chiesa, R. A. Popa: Oblix: An Efficient Oblivious Search Index. S&P 2018.

[MSW+14] Y-A. de Montjoye, E. Shmueli, SS. Wang, AS. Pentland: OpenPDS: Protecting the Privacy of Metadata through SafeAnswers. PLoS ONE 9(7) 2014

[MZC+16] R. Mortier, J. Zhao, J. Crowcroft, L. Wang, Q. Li, H. Haddadi, Y. Amar, A. Crabtree, J. Colley, T. Lodge, T. Brown, D. McAuley, C. Greenhalgh: Personal Data Management with the Databox: What's Inside the Box? ACM CoNEXT Cloud-Assisted Networking workshop, 2016

[OCF+15] O. Ohrimenko, M. Costa, C. Fournet, C. Gkantsidis, M. Kohlweiss, D.Sharma: Observing and Preventing Leakage in MapReduce. CCS 2015.

# References (4)


[OSF+16] O. Ohrimenko, F. Schuster, C. Fournet, A. Mehta, S. Nowozin, K. Vaswani, M. Costa: Oblivious Multi-Party Machine Learning on Trusted Processors. USENIX Security 2016.

[PGF+17] R. Pires, D. Gavril, P. Felber, E. Onica, M. Pasin: A lightweight MapReduce framework for secure processing with SGX. CCGrid 2017

[PVC18] C. Priebe, K. Vaswani, M. Costa: EnclaveDB: A Secure Database Using SGX. IEEE Symposium on Security and Privacy 2018: 264-278

[RHM19] L. Roche, J. M. Hendrickx, Y-A. de Montjoye: Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 2019

[SCF+15] F. Schuster, M. Costa, C. Fournet, C. Gkantsidis, M. Peinado, G. Mainar-Ruiz, M. Russinovich: VC3: Trustworthy Data Analytics in the Cloud Using SGX. S&P 2015

[TAP17] P. Tran-Van, N. Anciaux, P. Pucheral: SWYSWYK: A Privacy-by-Design Paradigm for Personal Information Management Systems. ISD 2017

[TCL+19] Y. Tang, J. Chen, K. Li, J. Xu, Q. Zhang: Authenticated Key-Value Stores with Hardware Enclaves. CoRR abs/1904.12068 (2019)

[WAK18] N. Weichbrodt, P.-L. Aublin, R. Kapitza: SGX-perf: A Performance Analysis Tool for Intel SGX Enclaves. Middleware 2018

[ZDB+17] W. Zheng, A. Dave, J. G. Beekman, R. A. Popa, J. E. Gonzalez, I. Stoica. Opaque: An oblivious and encrypted distributed analytics platform. NSDI 2017