Engineering privacy-preserving systems: Two case studies

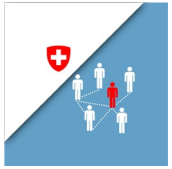
Prof. Carmela Troncoso

SPRING Lab
EPFL

8.09.2020

École polytechnique fédérale de Lausanne

And in reality…?

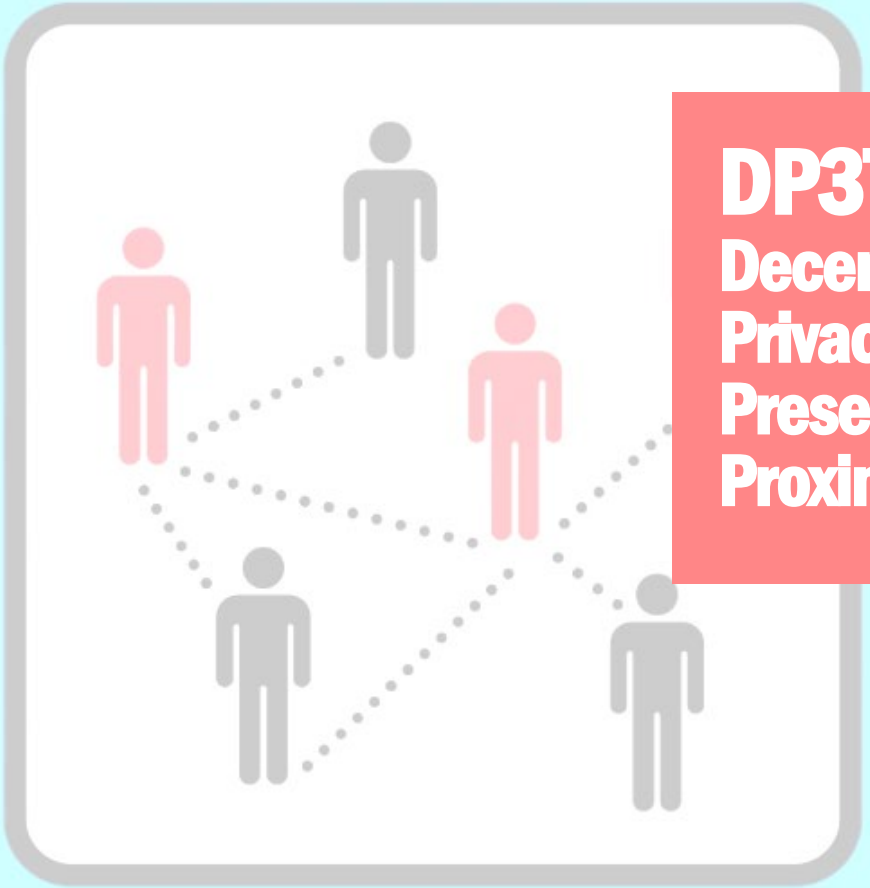


Contact Tracing App

- Design under time pressure
- Design under platform constraints
- Scale to billion of users
- Quasi real time

Decentralized Search Engine

- Loose time constraints
- Blank slate for design
- Hundred to thousand users
- Journalists are patient ☺



DP3T
Decentralized Privacy Preserving Proximity Tracing

A collaborative (continued) sprint

March 2020 - **Start**

April 2020 – **GAEN is announced**

May 2020 – **Final version DP3T**

June 2020 – **Pilot SwissCovid (& other EU apps)**

July 2020 – **SwissCovid launch**

August / Sept 2020 – **Towards international interoperability**

Decentralized Privacy-Preserving Proximity Tracing

Version: 25 May 2020.

Contact the first author for the latest version.

EPFL: Prof. Carmela Troncoso, Prof. Mathias Payer, Prof. Jean-Pierre Hubaux, Prof. Marcel Salathé, Prof. James Larus, Prof. Edouard Bugnion, Dr. Wouter Lueks, Theresa Stadler, Dr. Apostolos Pyrgelis, Dr. Daniele Antonioli, Ludovic Barman, Sylvain Chatel

ETHZ: Prof. Kenneth Paterson, Prof. Srdjan Čapkun, Prof. David Basin, Dr. Jan Beutel, Dr. Dennis Jackson, Dr. Marc Roeschlin, Patrick Leu

KU Leuven: Prof. Bart Preneel, Prof. Nigel Smart, Dr. Aysajan Abidin

TU Delft: Prof. Seda Gürses

University College London: Dr. Michael Veale

CISPA: Prof. Cas Cremers, Prof. Michael Backes, Dr. Nils Ole Tippenhauer

University of Oxford: Dr. Reuben Binns

University of Torino / ISI Foundation: Prof. Ciro Cattuto

Aix Marseille Univ, Université de Toulon, CNRS, CPT: Dr. Alain Barrat

IMDEA Software Institute: Prof. Dario Fiore

INESC TEC: Prof. Manuel Barbosa (FCUP), Prof. Rui Oliveira (UMinho), Prof. José Pereira (UMinho)

First: A clear goal

- **complement** manual contact tracing (identification COVID19+ patients' contacts for quarantining)
- in a **timely, efficient, and scalable** manner

- **notify** users that have been exposed to COVID19 and they are at risk of infection

Seda Gurses, Carmela Troncoso, Claudia Diaz. Engineering Privacy by Design.Computers, Privacy & Data Protection. 2011

Second: Security and Privacy Requirements

- Hide users identity, location, and behavior (social graph)
- Hide COVID+ users and contacts with COVID+ users
- Prevent false alarms
- Prevent Denial of Service

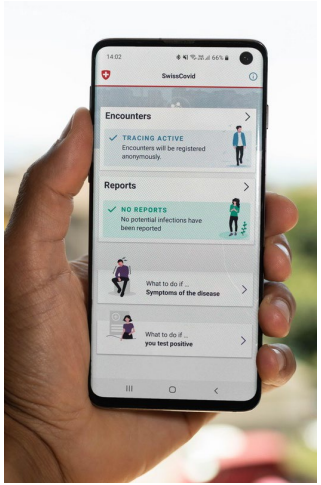
Seda Gurses, Carmela Troncoso, Claudia Diaz. Engineering Privacy by Design.Computers, Privacy & Data Protection. 20

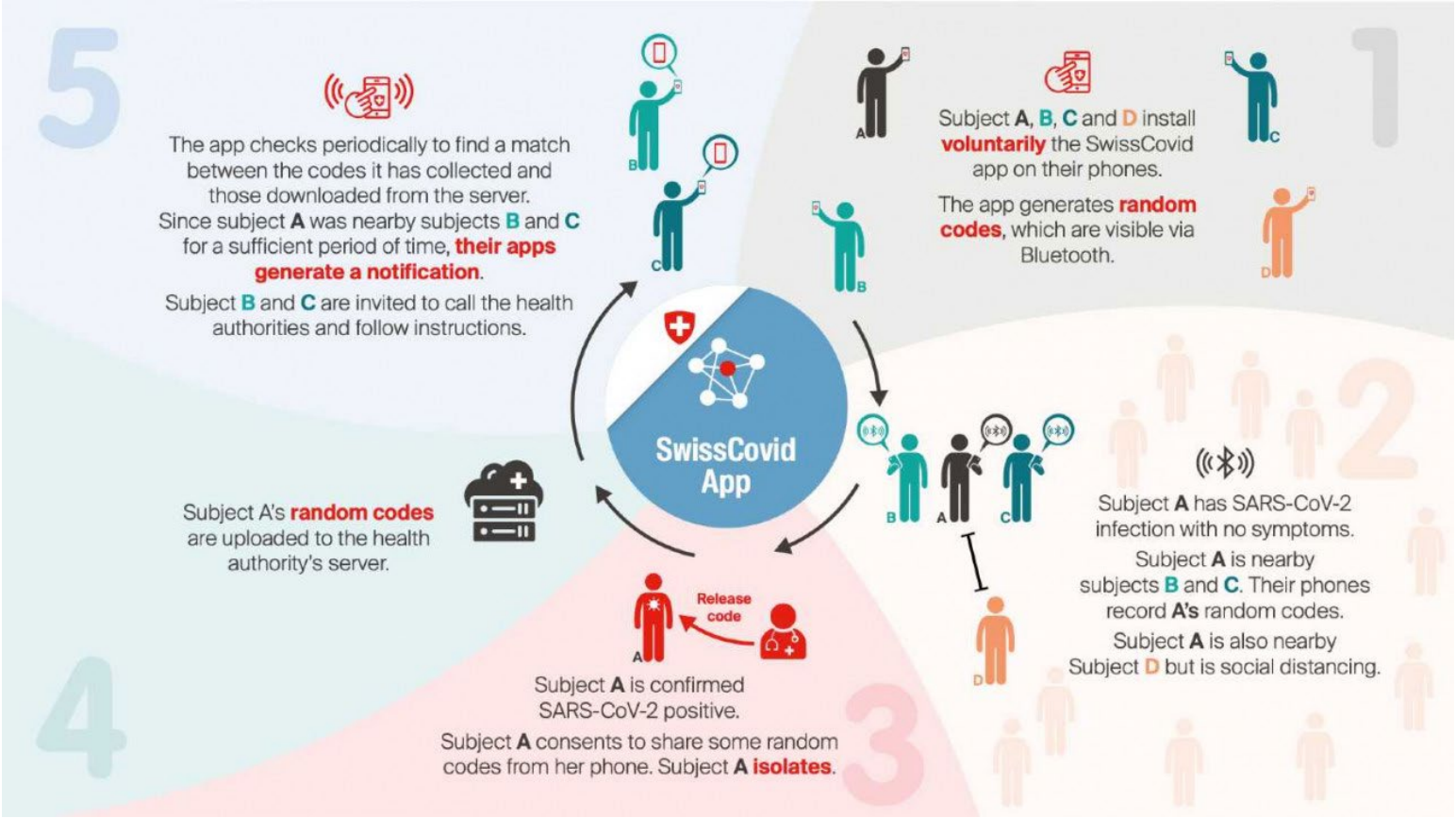
And… reality

- Design under pressure!

  (First deadline was 3 weeks!)
  - Follow the KISS principle
    needed fast, robust verification
  - Use existing infrastructure
    - Mobile platforms
    - Bluetooth technology

- High scalability and reliability
  - Avoid experimental / new technologies




5
The app checks periodically to find a match between the codes it has collected and those downloaded from the server.
Since subject A was nearby subjects B and C for a sufficient period of time, their apps generate a notification.
Subject B and C are invited to call the health authorities and follow instructions.
1
Subject A, B, C and D install voluntarily the SwissCovid app on their phones.
The app generates random codes, which are visible via Bluetooth.
SwissCovid App
Subject A's random codes are uploaded to the health authority's server.
2
Subject A has SARS-CoV-2 infection with no symptoms.
Subject A is nearby subjects B and C. Their phones record A's random codes.
Subject A is also nearby Subject D but is social distancing.
4
Release code
Subject A is confirmed SARS-CoV-2 positive.
Subject A consents to share some random codes from her phone. Subject A isolates.
3

Result: Security and privacy by design

- **Only** information that ever leaves the phone are the **random numbers** (no identity, no location, no information about others) **broadcasted** during the contagious period
- **No** information available for abuse
- System **sunsets-by-design**

- Attacks inherent to the underlying technology can't be avoided
  - Other attacks, non-trivial cost

Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems

The DP-3T Project

21 April 2020

The basic idea behind digital proximity tracing through mobile applications is to use Bluetooth Low Energy (BLE) signals to estimate physical proximity between two smartphones. The only functionality that such an app needs to provide is to inform the contacts of an infected person that they might have been exposed to the virus through a close-range physical contact. The system **does not** need to reveal to anyone **who** the potential contagious contact was **with,** or **when** and **where** it happened.

This document summarises the findings of an in-depth privacy and security analysis of digital proximity tracing systems that our team has conducted over the past few weeks. It lists the risks inherent to any digital proximity tracing system, risks inherent to systems based on BLE handshakes between personal smartphones, and additional risks of proposed design variants of the latter. The risk analysis in this document was primarily conducted by the DP-3T team, but it is also informed by online discussions on the project's GitHub repository, ePrint reports, as well as exchanges via email and other platforms. We are grateful for all of this assistance and welcome suggestions about risks we missed.

https://github.com/DP-3T/documents/blob/master/Security%20analysis/Privacy%20and%20Security%20Attacks%20on%20Digital%20Proximity%20Tracing%20Systems.pdf

And if you want an app… Google & Apple will play a role

“When we build something we have to pick an architecture that works. And it has to work globally, for all countries around the world. And when we did the analysis and looked at different approaches we were very heavily inspired by the **DP-3T** group and their approach — and that’s what we have adopted as a solution. We think that gives the best privacy preserving aspects of the contacts tracing service.”

Dave Burke, VP of Android, Google, ***24.4.2020***

https://techcrunch.com/2020/04/24/apple-and-google-update-joint-coronavirus-tracing-tech-to-improve-user-privacy-and-developer-flexibility/

“Contact tracing can help slow the spread of COVID-19 and can be done without compromising user privacy. We’re working with @sundarpichai & @Google to help health officials harness Bluetooth technology in a way that also respects transparency & consent.”

Time Cook, CEO, Apple, ***10.4.2020*** *on Twitter*

The protocol…
A small piece in the puzzle

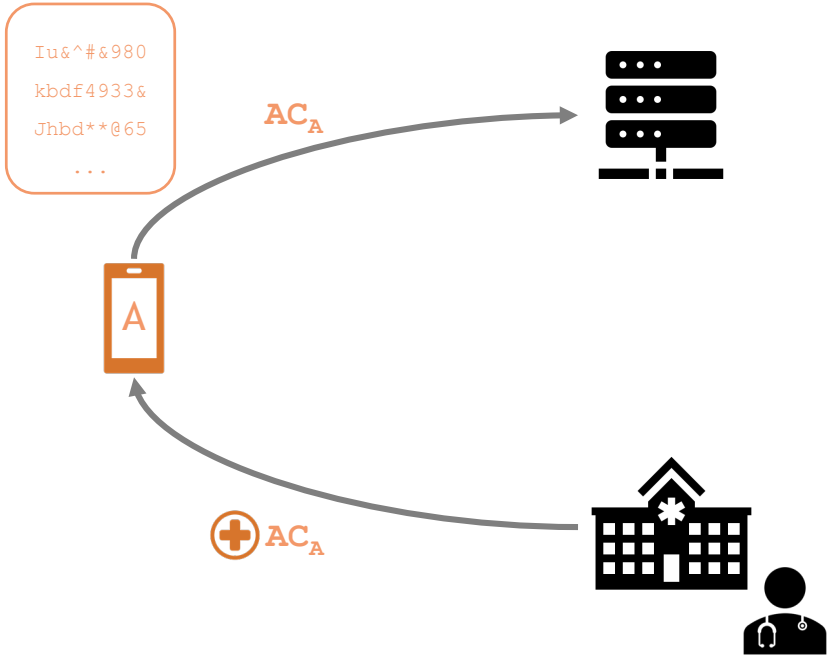
Interdisciplinary team (30+ researchers, 10 countries)

Privacy, Systems, Cryptography, Wireless security, SW Security, Requirements engineering, Epidemiologists, Ethicists, Law

Carmela Troncoso

Connecting to the Health system

My slide in non-technical talks

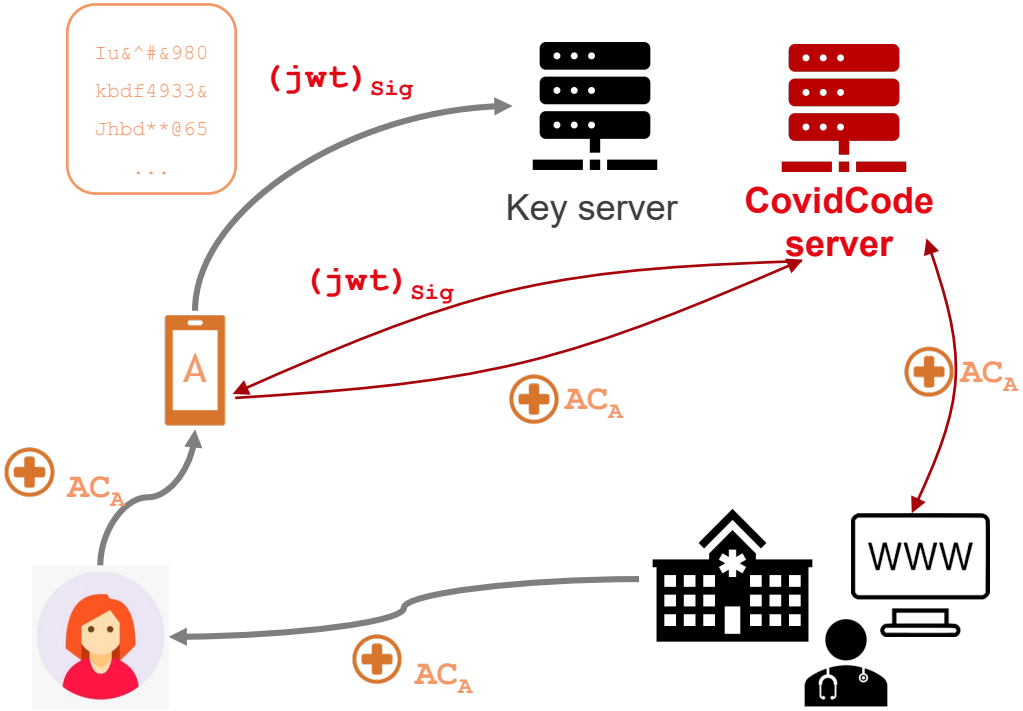


When a user is diagnosed positive, if they consent, they upload their keys (their numbers)

These numbers…

- Are not related to **A**'s identity
- Are not related to the locations **A** visited
- Are not related to other people **A** has interacted with or has seen

Connecting to the Health system
Reality

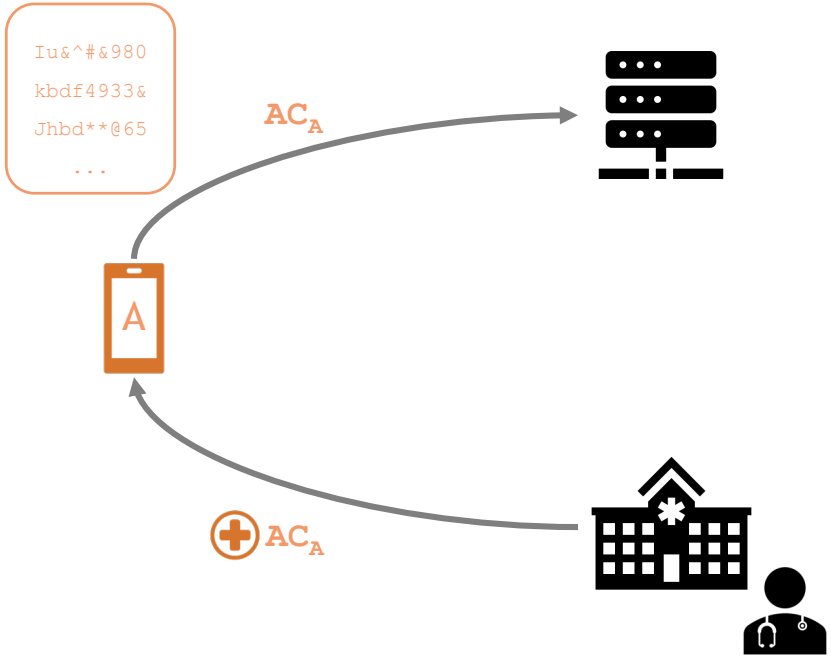


When a user is diagnosed positive, if they consent, they upload their keys (their numbers)

These numbers…

- Are not related to **A**'s identity
- Are not related to the locations **A** visited
- Are not related to other people **A** has interacted with or has seen

Connecting to the Health system
Does it matter for privacy?



Uploading keys => the user is COVID+

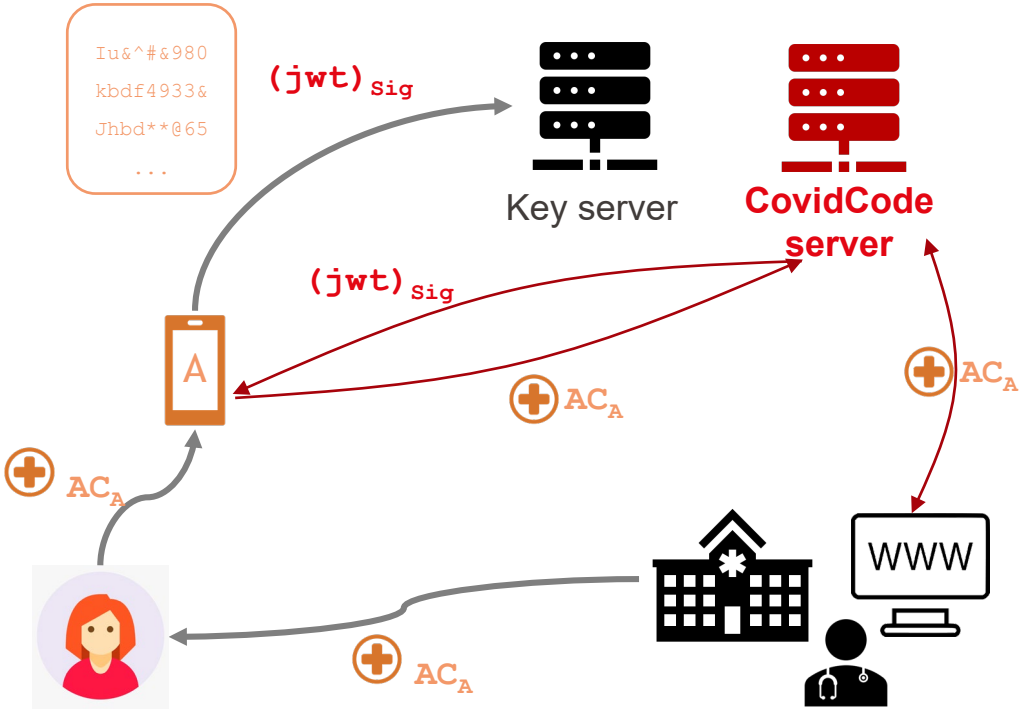
DP3T design paper

The pattern associated with the upload of identifiers to the server would reveal the COVID-19 positive status of users to network eavesdroppers (ISP or curious WiFi provider) and tech-savvy adversaries. If these adversaries can bind the observed IP address to a more stable identifier such as an ISP subscription number, then they can de-anonymize the confirmed positive cases. This can be mitigated by using dummy uploads. These

- https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf

Carmela Troncoso

Connecting to the Health system
Does it matter for privacy?



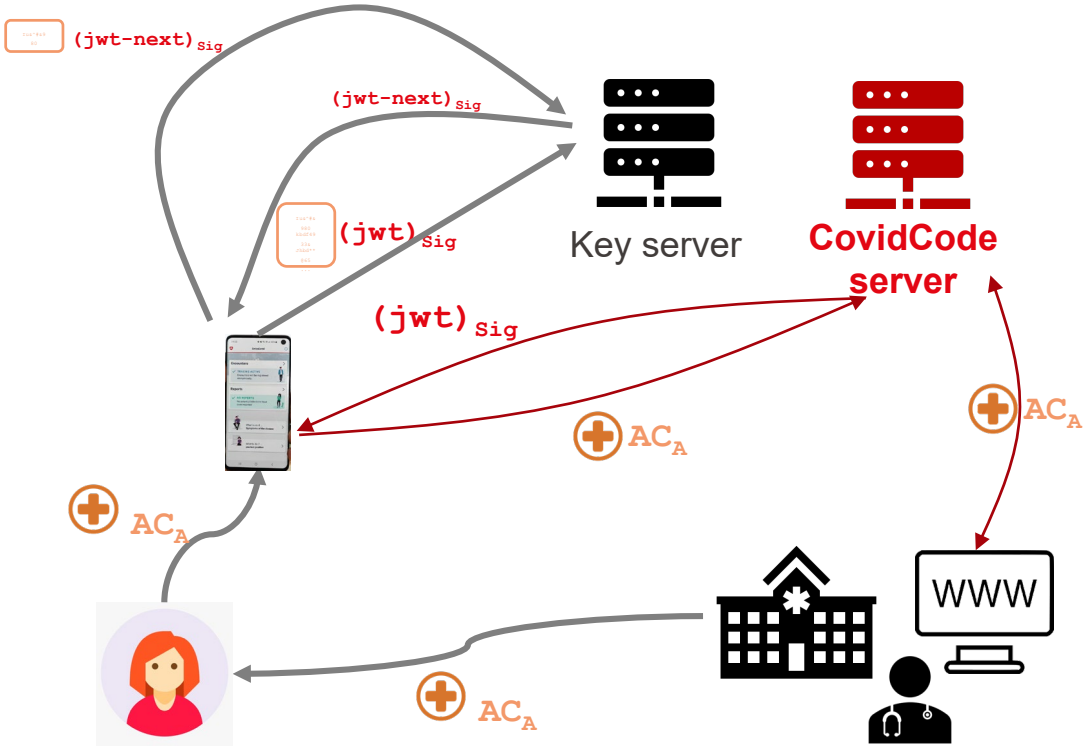
Dummy traffic design

- How to schedule dummies?
  - Battery consumption
  - Bandwidth consumption
  - Real behavior???
- Plausible deniability:
  - There is no anonymity possible
  - Real uploads must happen
  - Constant size & time operations!
- Include the authentication step!

- https://github.com/DP-3T/documents/blob/master/DP3T%20-%20Best%20Practices%20for%20Operation%20Security%20in%20Proximity%20Tracing.pdf

Carmela Troncoso

Connecting to the Health system
Reality gets worse

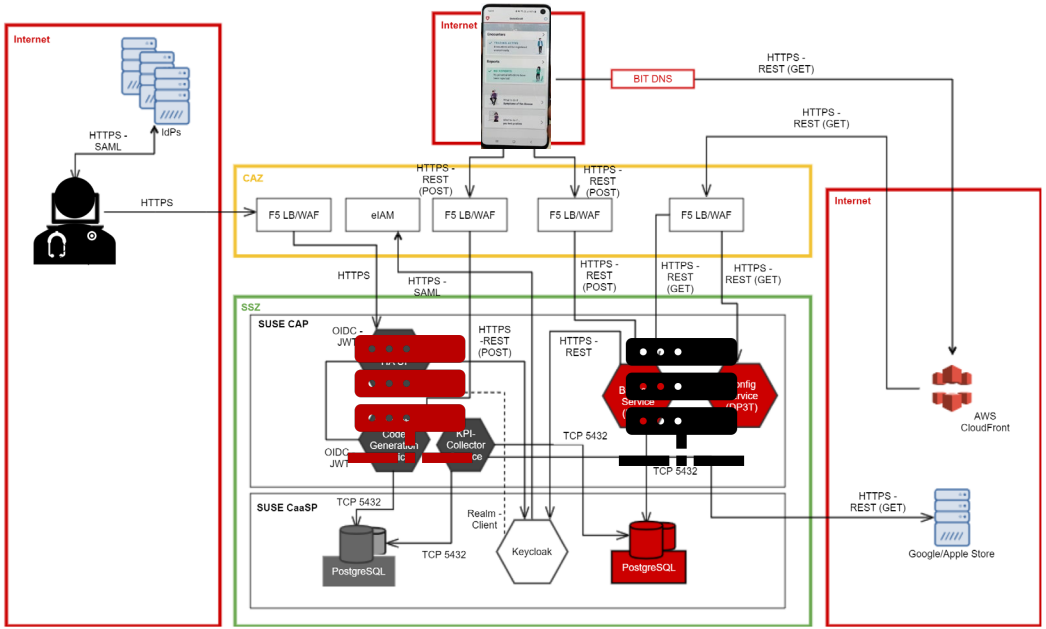


GAEN < v1.5… has a "security" feature. Keys are only returned after they expire

- Rock & hard place
  - Delay all keys one day
  - **Have a second upload**
    - to be mimicked by dummies
- Processes in the phone can't wake up often, and they have little processing time…
  - Sometimes they do not wake up
  - Anecdote: retrieving configuration when the user opens the app… Another leak! (and another action to mimic)

Carmela Troncoso

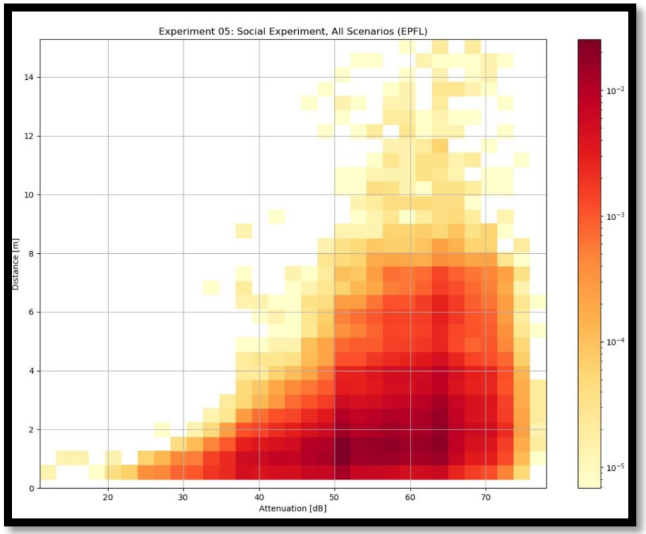
Connecting to the Health system
Reality gets even worse



Servers live in an infrastructure with other security mechanisms

- DoS prevention, load balancing
- Ensure that
  - **CovidCode** server logs do not match **Key Server** logs
  - Cloud logs do not match **Key server** logs
- **Key server** logs little and **very** coarse

More on reality
beyond privacy engineering



Bluetooth configuration
(on an ever-changing API)

Ordonnance sur le système de traçage de proximité pour le coronavirus SARS-CoV-2

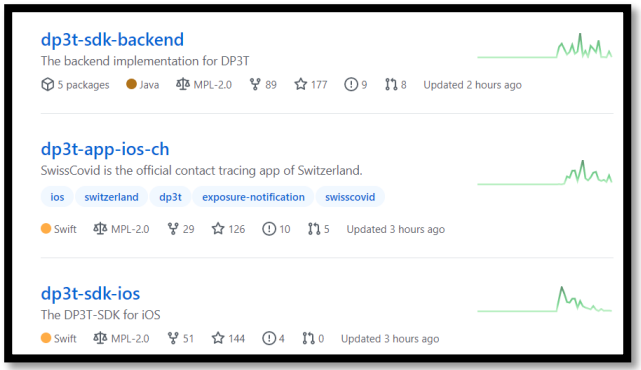
(OSTP)

du 24 juin 2020 (Etat le 25 juin 2020)

Le Conseil fédéral suisse,

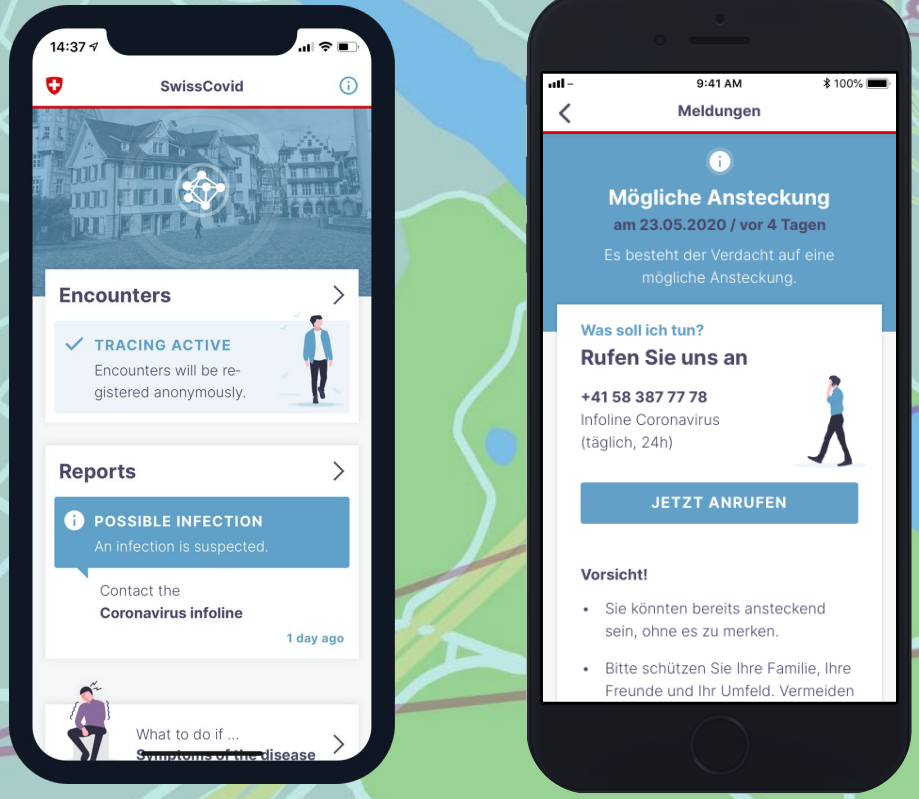
vu l'art. 60a, al. 7, de la loi du 28 septembre 2012 sur les épidémies (LEp)[1],

Legal framework
Helps with quarantine, testing, discrimination

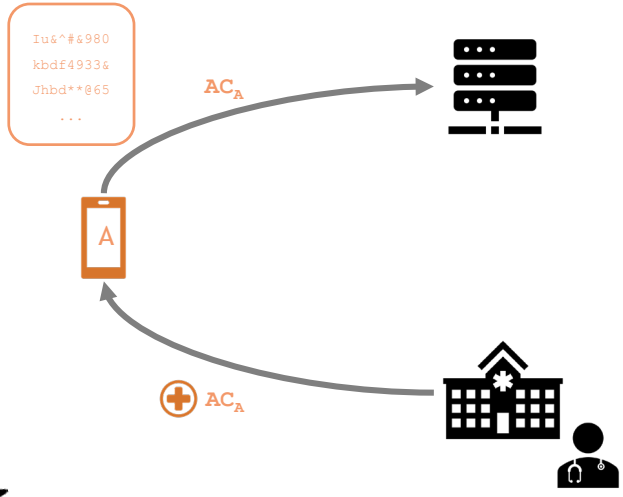


Ever-changing code
API changes, functionality, UX,…

And all this (and more) is SwissCovid



Operation requires monitoring but... privacy



DP3T design paper

The pattern associated with the upload of identifiers to the server would reveal the COVID-19 positive status of users to network eavesdroppers (ISP or curious WiFi provider) and tech-savvy adversaries. If these adversaries can bind the observed IP address to a more stable identifier such as an ISP subscription number, then they can de-anonymize the confirmed positive cases. This can be mitigated by using dummy uploads. These

- https://www.experimental.bfs.admin.ch/bfsstatic/dam/assets/13667538/master

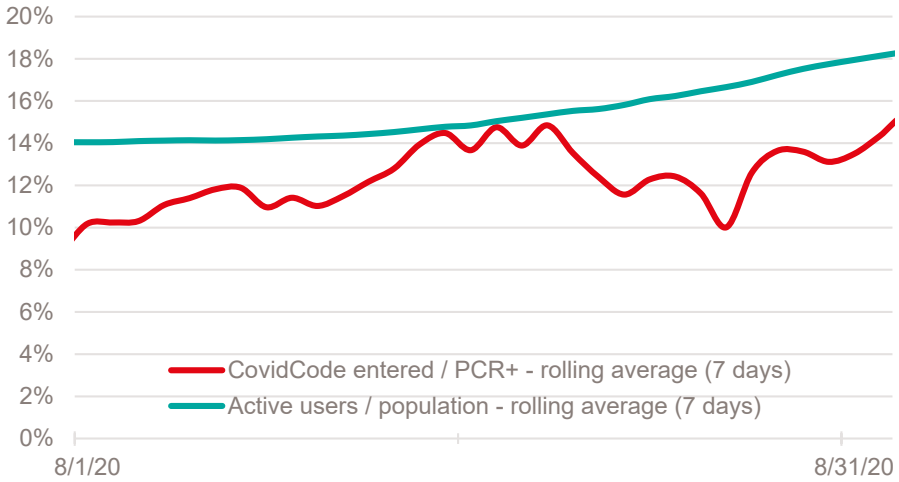
Monitoring privacy-preserving is hard. Only partial observation of actions in the system.

We can count

- Issued and uploaded CovidCodes
- Active apps
  - Dummy uploads: 1 every 5 days on average
  - Count number of uploads * 5

Other monitoring must be **offline**

Promising results and Growing Momentum



Data from August 2020 (1mo)

- 1.6 Million active users (~18%)
- 1054 users uploaded their keys (12.4% of PCR in Switzerland)
- Confirmed in the field
  - 26 positive cases found that would have not been tested without the app!
  - First indications of effectiveness, speed, and complementarity to classic contact tracing

https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.html
https://github.com/digitalepidemiologylab/swisscovid_efficacy/blob/master/SwissCovid_efficacy_MS.pdf

Broad European Adoption

Proximity Tracing beyond borders

- Cross-border interoperability
  - Users only need to consent to share **their** keys
- Allows national apps to work when roaming
- Two basic use cases:
  - Subscribe to foreign feeds when roaming (and the 10 days that follow)
  - Add keys to another country's set if contagious in that country
- Will be operated by EC e-Health Network, starting in October.

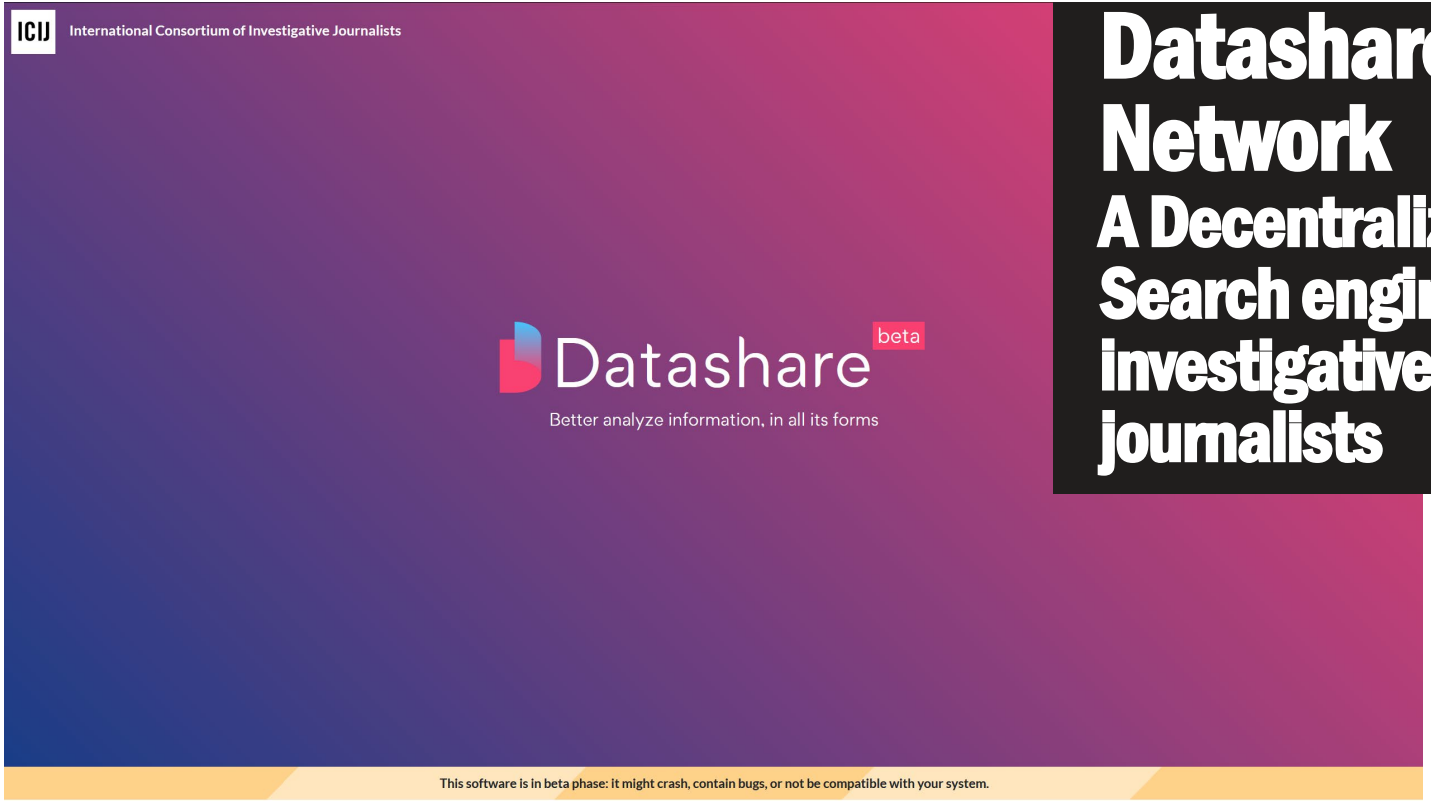
- **A NEW HELL FOR PRIVACY!**
  - **Hiding travelers**
  - **Hiding destinations**

Summary

- **First privacy-by-design product developed at large scale** with collaboration of key players in the mobile industry

- Key lessons
  - Integration in Health System is key (and hard)
  - Privacy engineering in an agile/service world is exhausting

- Steps ahead
  - Effectiveness indicators
  - Interoperability across borders



Datashare Network

A Decentralized Search engine for investigative journalists

EPFL

ICIJ INTERNATIONAL CONSORTIUM of INVESTIGATIVE JOURNALISTS

INVESTIGATIONS INSIDE ICIJ DATA JOURNALISTS ABOUT DONATE FOLLOW

Journalists

The International Consortium of Investigative Journalists is a global network of 267 investigative journalists in 100 countries who collaborate on in-depth investigative stories.

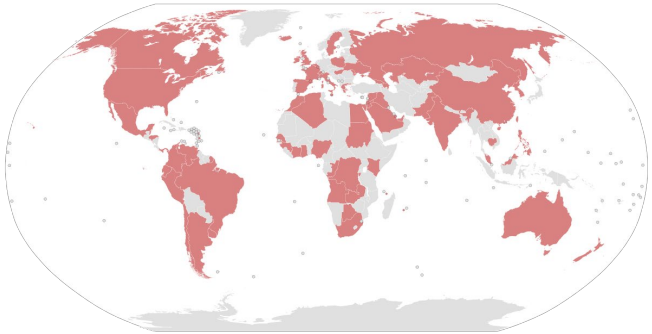
Site: icij.org

An ICIJ Investigation

The Panama Papers: Exposing the Rogue Offshore Finance Industry

A giant leak of more than 11.5 million financial and legal records exposes a system that enables crime, corruption and wrongdoing, hidden by secretive offshore companies.

11.5 Million documents
(Centralized)



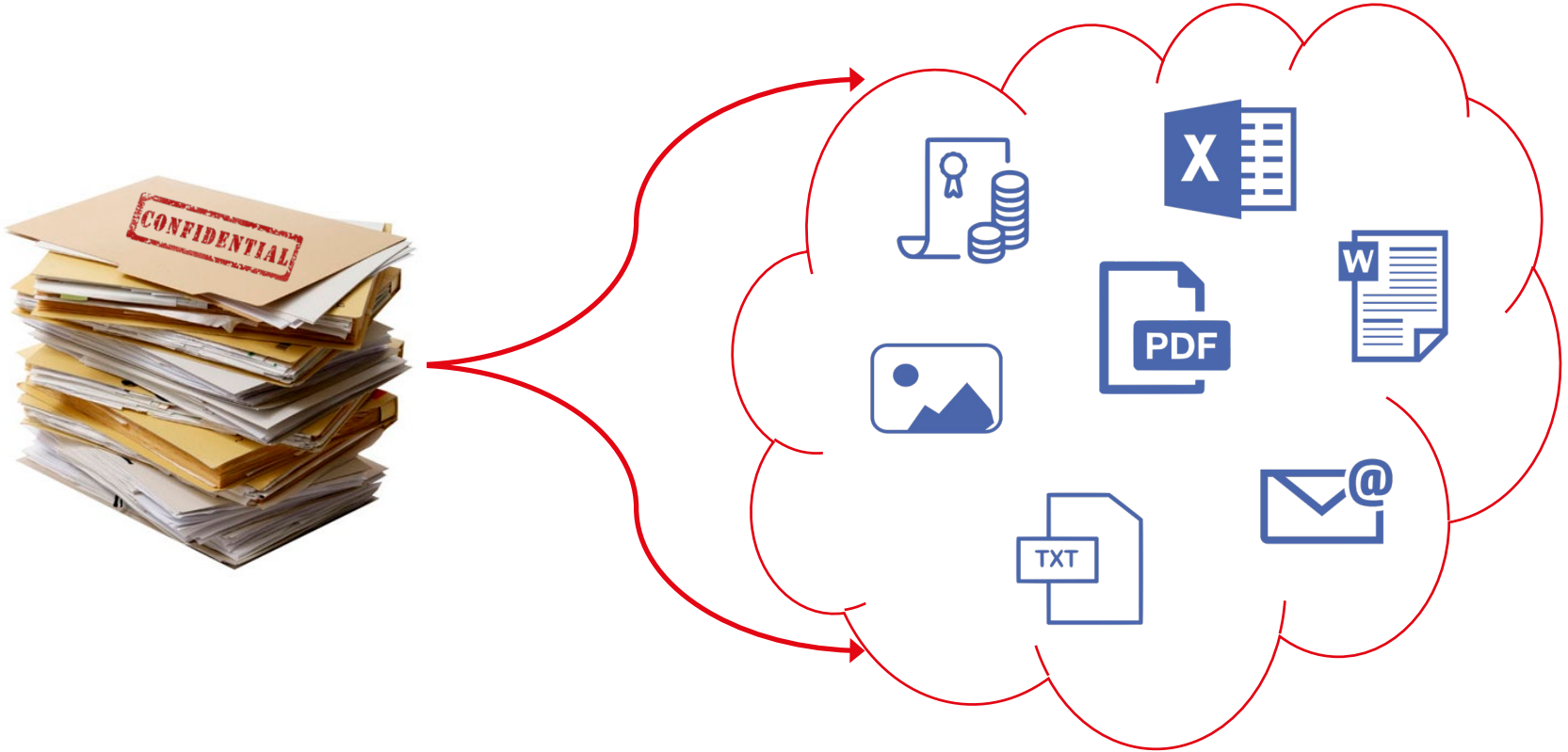
Journalists from over
80 countries

An ICIJ Investigation

The Panama Papers: Exposing the Rogue Offshore Finance Industry

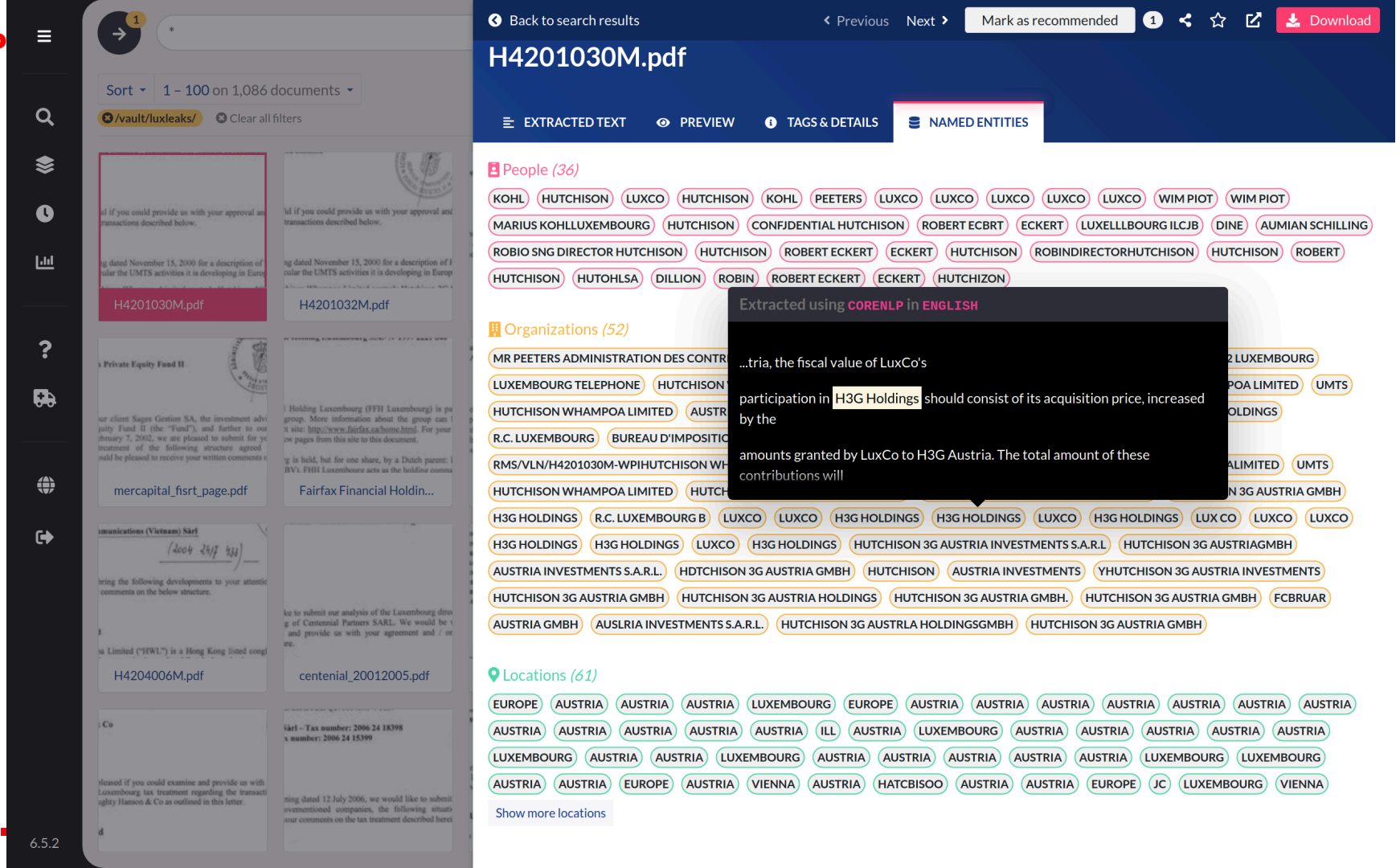
A giant leak of more than 11.5 million financial and legal records exposes a system that enables crime, corruption and wrongdoing, hidden by secretive offshore companies.

Digitalization



Better analyze information, in all its forms

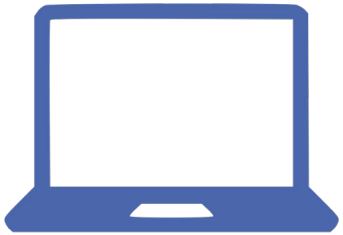
Site: datashare.icij.org


Back to search results
Previous
Next
Mark as recommended
Download
H4201030M.pdf
EXTRACTED TEXT
PREVIEW
TAGS & DETAILS
NAMED ENTITIES
People (36)
KOHL HUTCHISON LUXCO HUTCHISON KOHL PEETERS LUXCO LUXCO LUXCO LUXCO LUXCO WIM PIOT WIM PIOT
MARIUS KOHLLUXEMBOURG HUTCHISON CONFJDENTIAL HUTCHISON ROBERT ECBRT ECKERT LUXELLLBOURG ILCJB DINE AUMIAN SCHILLING
ROBIO SNG DIRECTOR HUTCHISON HUTCHISON ROBERT ECKERT ECKERT HUTCHISON ROBINDIRECTORHUTCHISON HUTCHISON ROBERT
HUTCHISON HUTOHLSA DILLION ROBIN ROBERT ECKERT ECKERT HUTCHIZON
Extracted using CORENLP in ENGLISH
...tria, the fiscal value of LuxCo's
participation in H3G Holdings should consist of its acquisition price, increased by the
amounts granted by LuxCo to H3G Austria. The total amount of these contributions will
Organizations (52)
MR PEETERS ADMINISTRATION DES CONTR
LUXEMBOURG TELEPHONE HUTCHISON
HUTCHISON WHAMPOA LIMITED AUSTR
R.C. LUXEMBOURG BUREAU D'IMPOSITIC
RMS/VLN/H4201030M-WPIHUTCHISON WH
HUTCHISON WHAMPOA LIMITED HUTCH
H3G HOLDINGS R.C. LUXEMBOURG B LUXCO LUXCO H3G HOLDINGS H3G HOLDINGS LUXCO H3G HOLDINGS LUX CO LUXCO LUXCO
H3G HOLDINGS H3G HOLDINGS LUXCO H3G HOLDINGS HUTCHISON 3G AUSTRIA INVESTMENTS S.A.R.L HUTCHISON 3G AUSTRIAGMBH
AUSTRIA INVESTMENTS S.A.R.L. HDTCHISON 3G AUSTRIA GMBH HUTCHISON AUSTRIA INVESTMENTS YHUTCHISON 3G AUSTRIA INVESTMENTS
HUTCHISON 3G AUSTRIA GMBH HUTCHISON 3G AUSTRIA HOLDINGS HUTCHISON 3G AUSTRIA GMBH. HUTCHISON 3G AUSTRIA GMBH FCBRUAR
AUSTRIA GMBH AUSLRIA INVESTMENTS S.A.R.L. HUTCHISON 3G AUSTRLA HOLDINGSGMBH HUTCHISON 3G AUSTRIA GMBH
Locations (61)
EUROPE AUSTRIA AUSTRIA AUSTRIA LUXEMBOURG EUROPE AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA
AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA ILL AUSTRIA LUXEMBOURG AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA
LUXEMBOURG AUSTRIA AUSTRIA LUXEMBOURG AUSTRIA AUSTRIA AUSTRIA AUSTRIA AUSTRIA LUXEMBOURG LUXEMBOURG
AUSTRIA AUSTRIA EUROPE AUSTRIA VIENNA AUSTRIA HATCBISOO AUSTRIA AUSTRIA EUROPE JC LUXEMBOURG VIENNA
Show more locations
Sort
1 – 100 on 1,086 documents
/vault/luxleaks/
Clear all filters
H4201030M.pdf
H4201032M.pdf
mercapital_fisrt_page.pdf
Fairfax Financial Holdin...
H4204006M.pdf
centenial_20012005.pdf
6.5.2

First:
A (not so)clear goal

Central

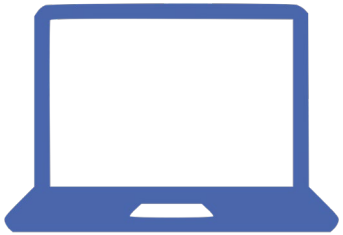
First:
A (not so)clear goal



Local

Central

First:
A (not so)clear goal



Local

Journalist ↔ Journalist

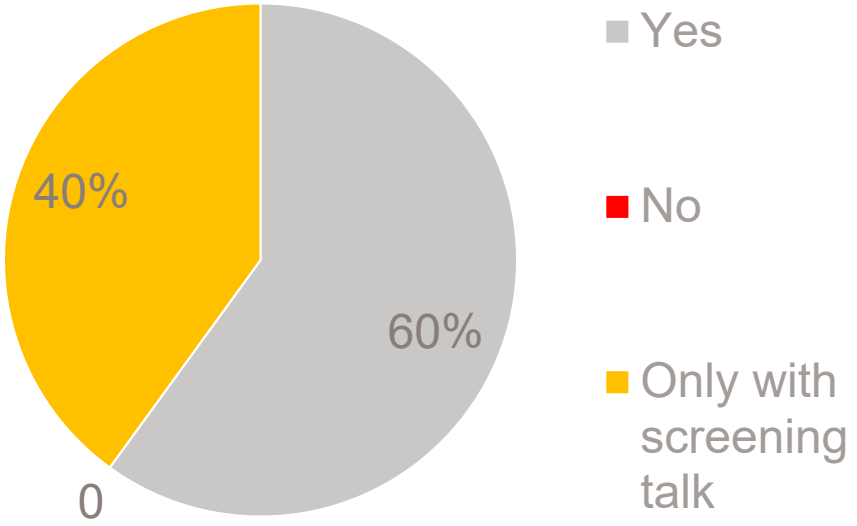
Central

Clarifying the goal

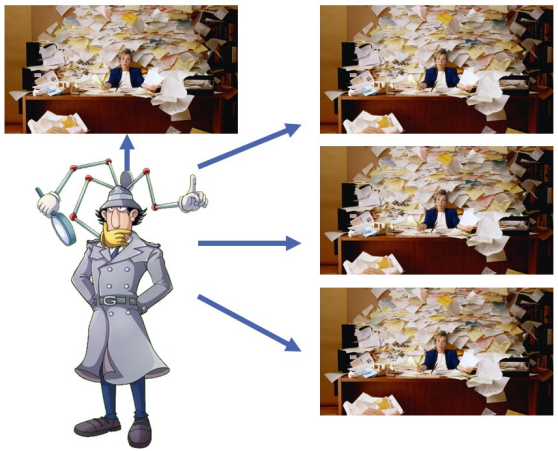
- ICIJ's survey among 70 members
  - Functionality
  - Resources
  - Concerns
- Weekly meetings during 1.5 years
  - Refinement
  - Negotiation

Survey: sharing

Are you willing to share your documents?

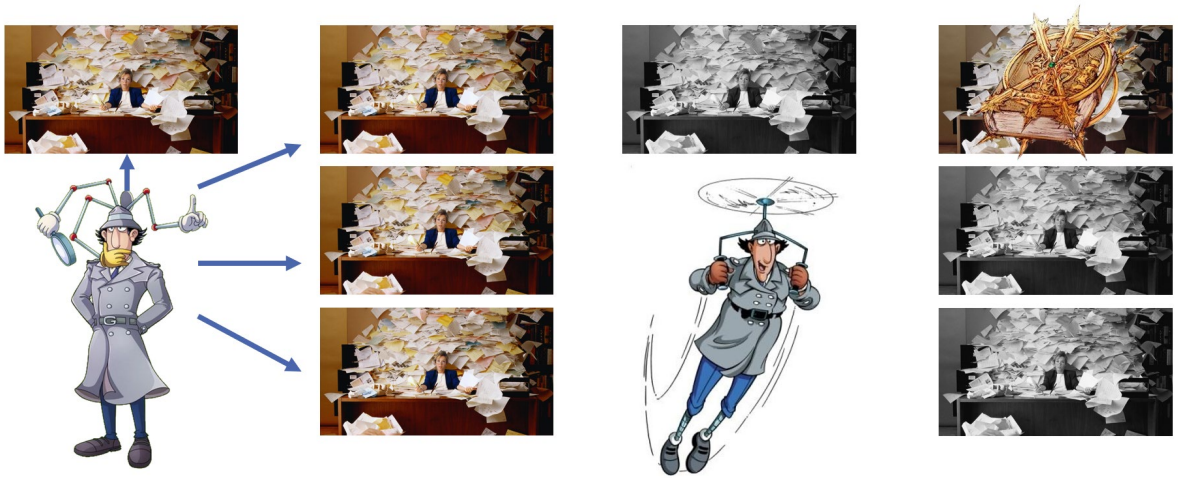


1. Privacy-preserving search



Search

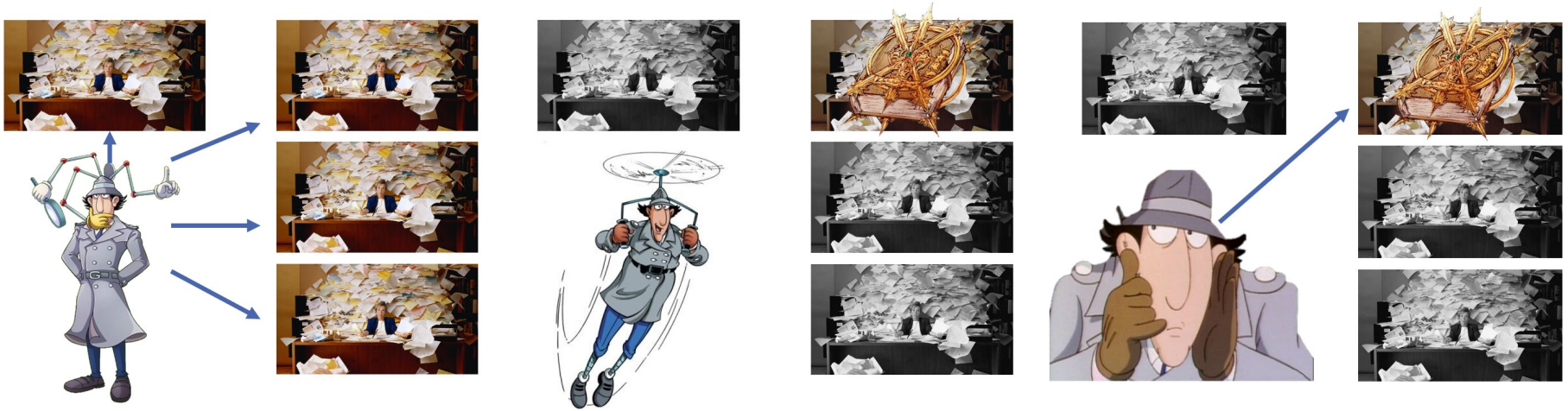
1. Privacy-preserving search



Search

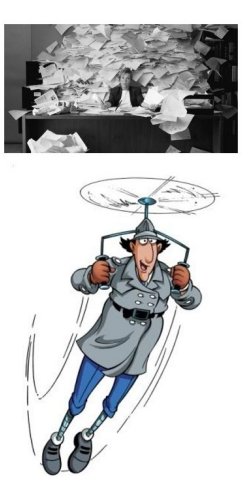
Find

2. Screening talk



Search Find Contact

Out of band collaboration/sharing



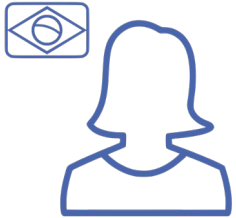
Find

Contact

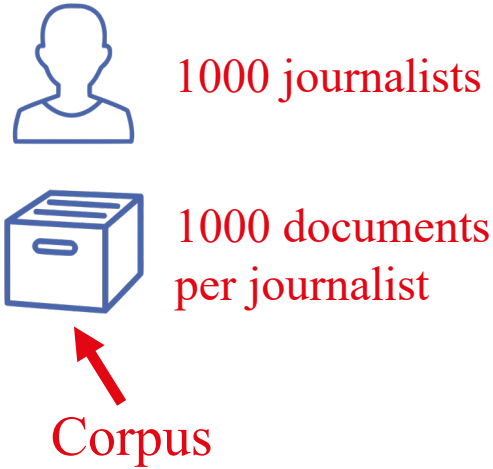
No retrieval

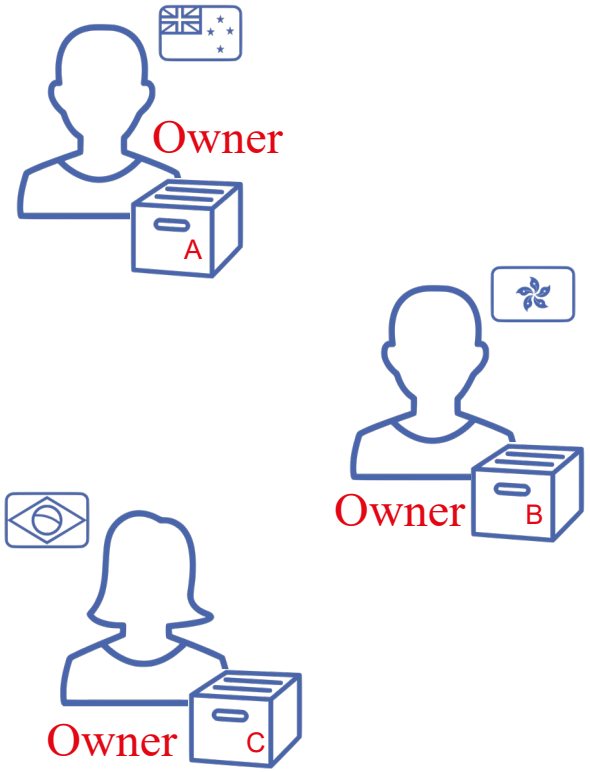
Setting

~~267~~ 1000 journalists



Setting


1000 journalists
1000 documents
per journalist
Corpus

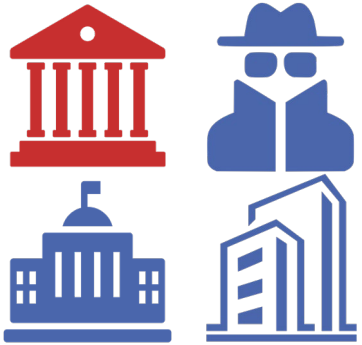

Owner
A
Owner
B
Owner
C

Second: Security and Privacy Requirements

Journalists

ICIJ

ICIJ

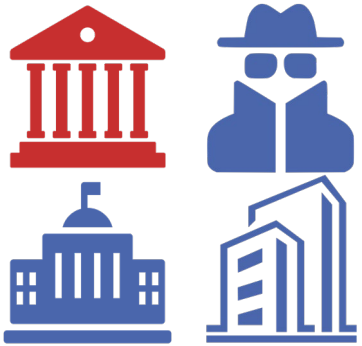


Third party

Second: Security and Privacy Requirements

Journalists

ICIJ

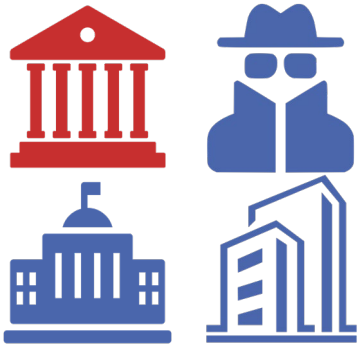


Third party

Second: Security and Privacy Requirements

Journalists

ICIJ

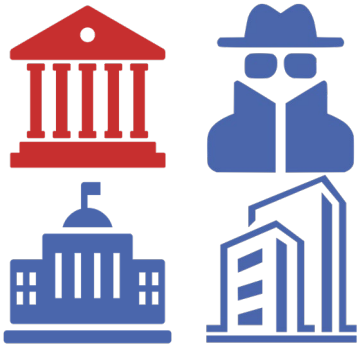


Third party

Second: Security and Privacy Requirements

Journalists

ICIJ



Third party

Second: Security and Privacy Requirements
Coerce or compromise

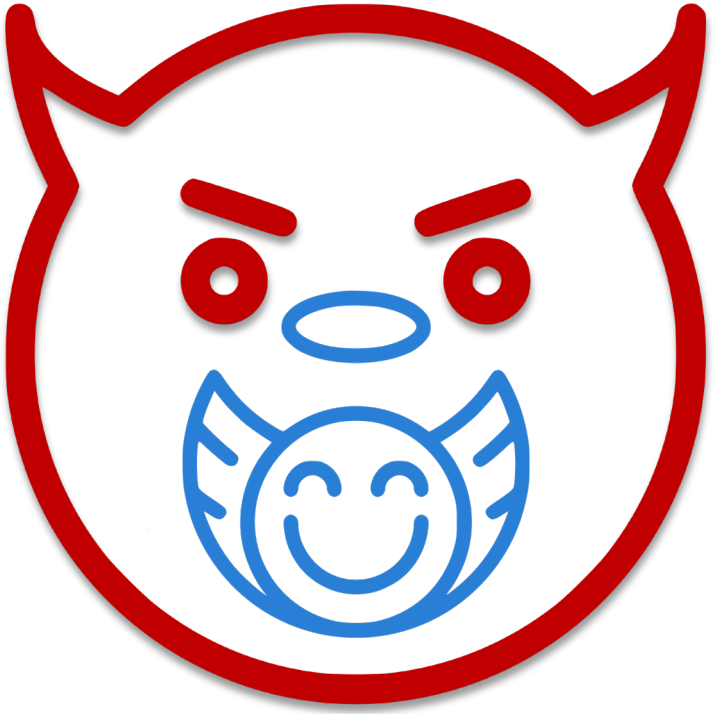
Hacking

Violence

Subpoena

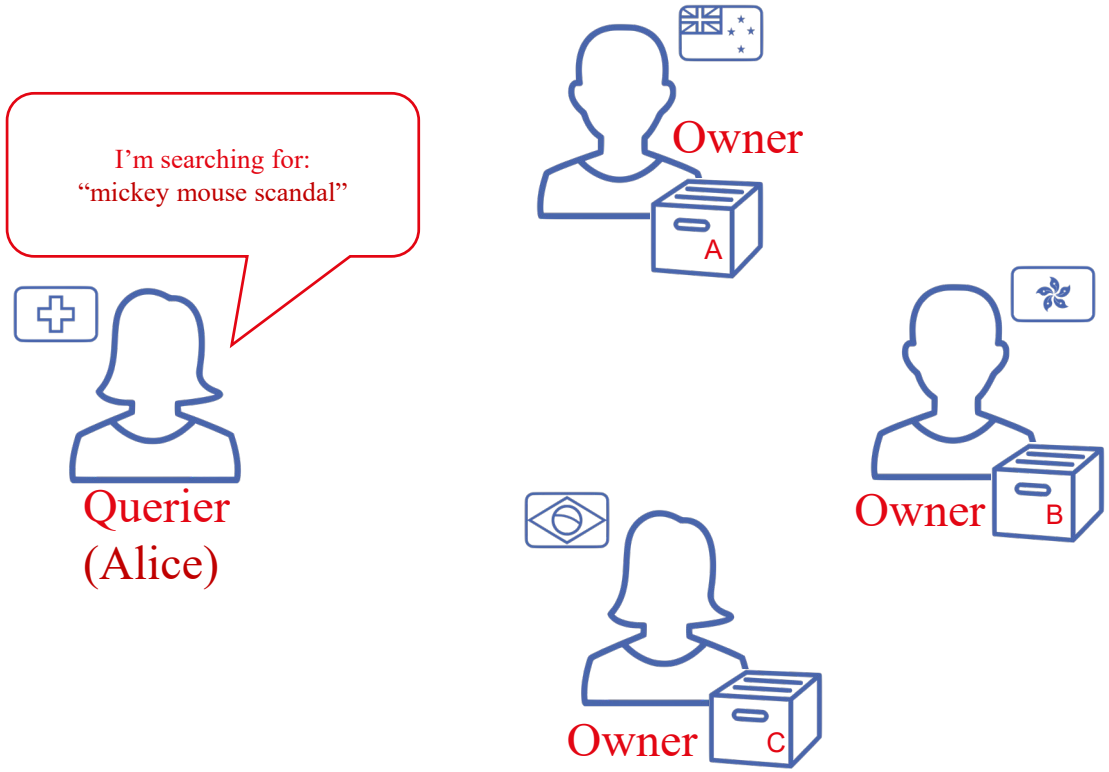
Second: Security and Privacy Requirements

ICIJ



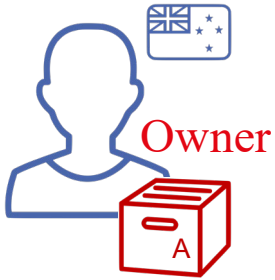
Privacy

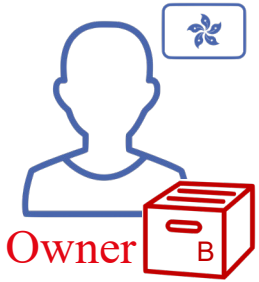
Privacy: querier

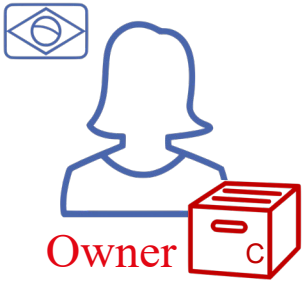


Protect:
- The query
- The querier's identity

Privacy: owner

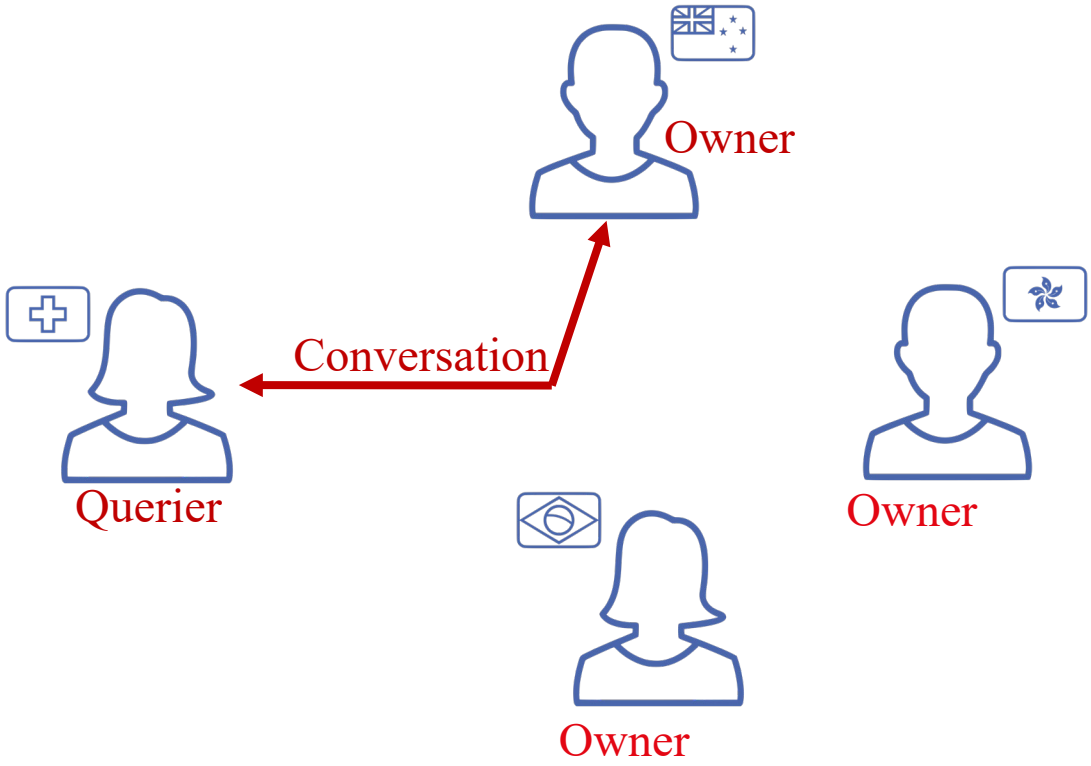






Protect:
- The corpus
- The owner's identity

Privacy: screening



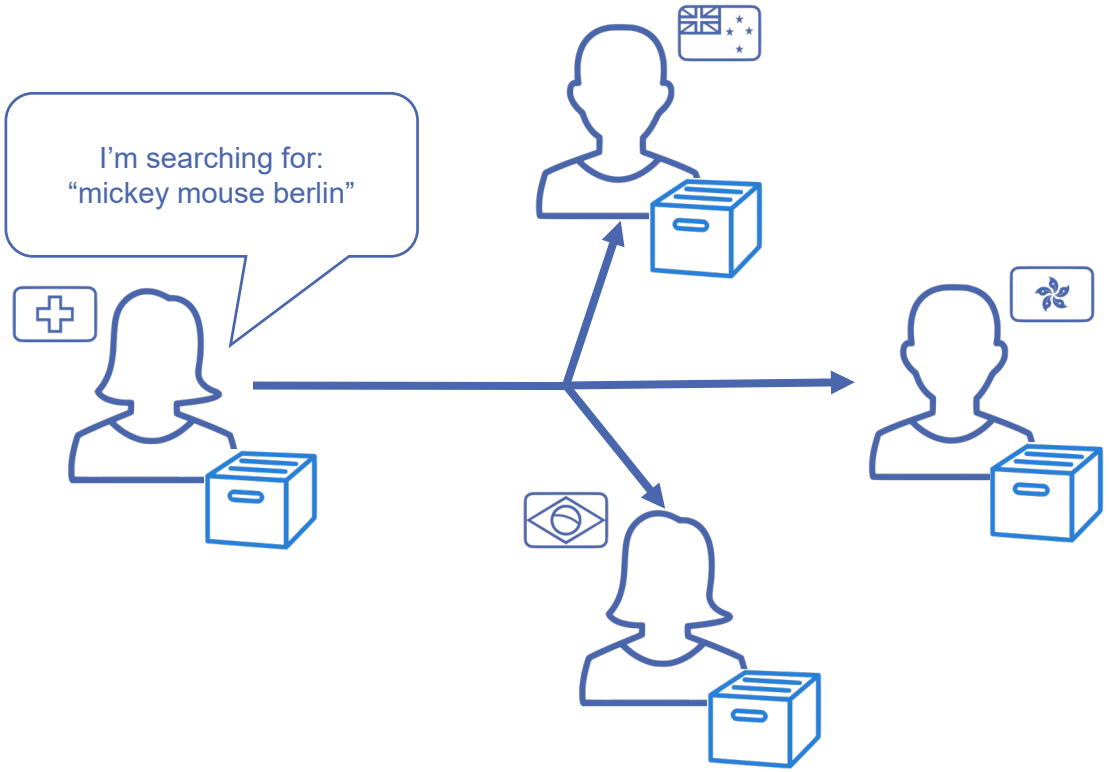
Protect:
- The content of conversation
- The existence of conversation

Datashare Network

Enable journalists to search on others' collections for keywords of interest.

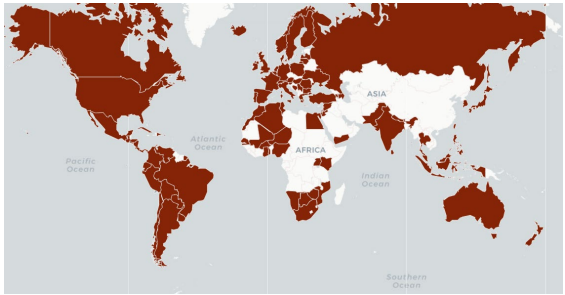
Protect journalists & sources.

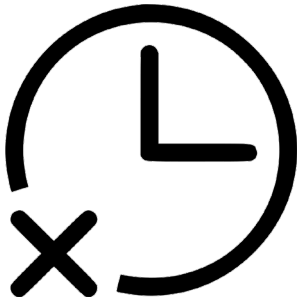
- Only ICIJ and associates can use the system.
- No one (journalists, ICIJ, others) can learn:
  who queries
  what is queried
- Journalists can anonymously converse with journalists that have matching documents.



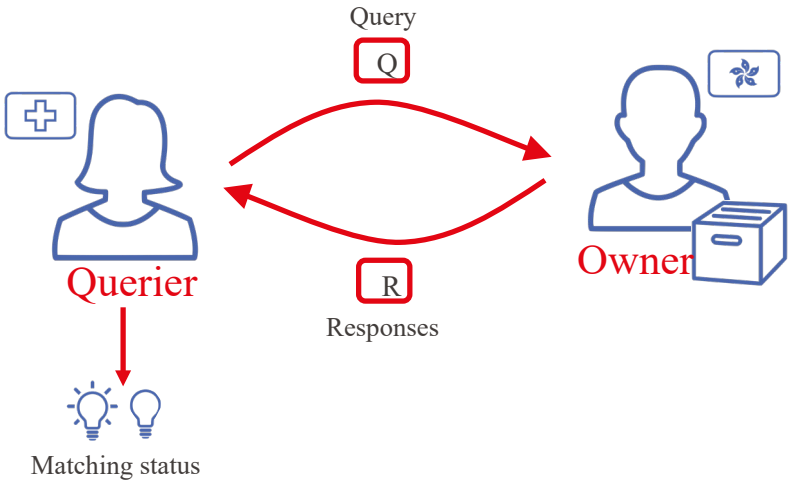
And… reality

- Asynchrony
- Scarce resources
  - Computation
  - Bandwidth
- But… no real time
  - Avoid experimental / new technologies





Privacy-preserving search

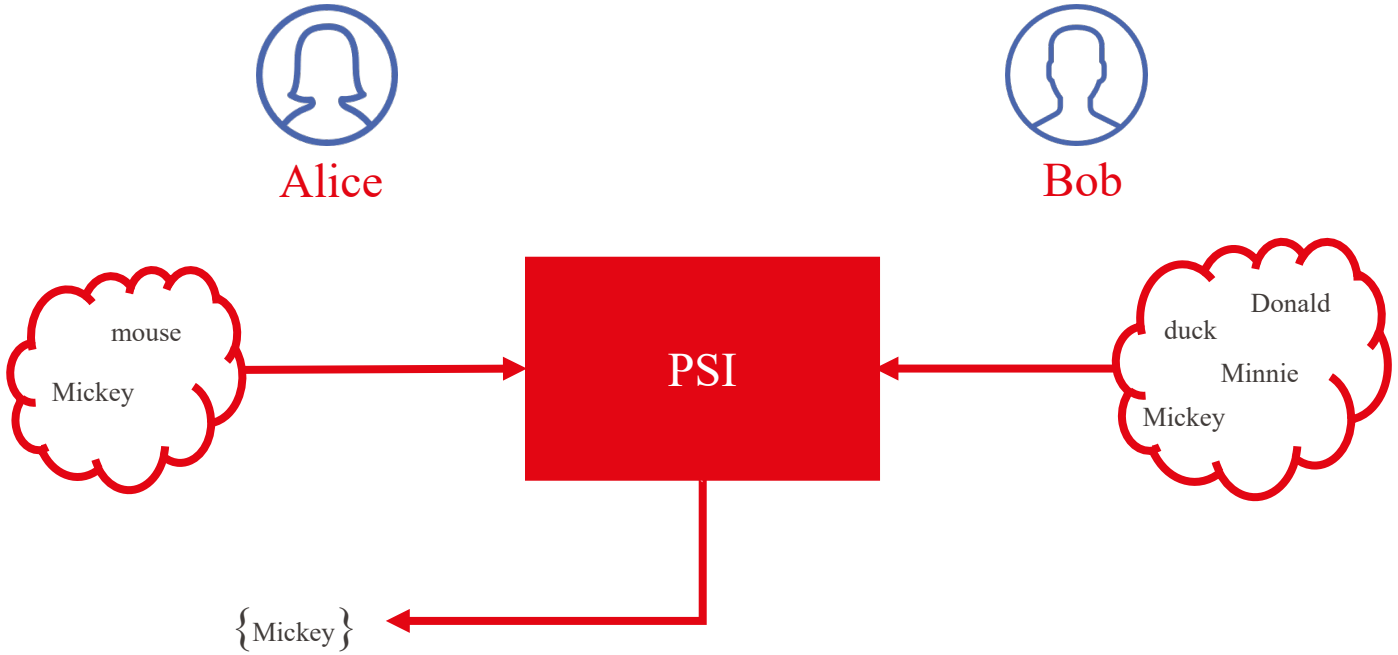


Protect:

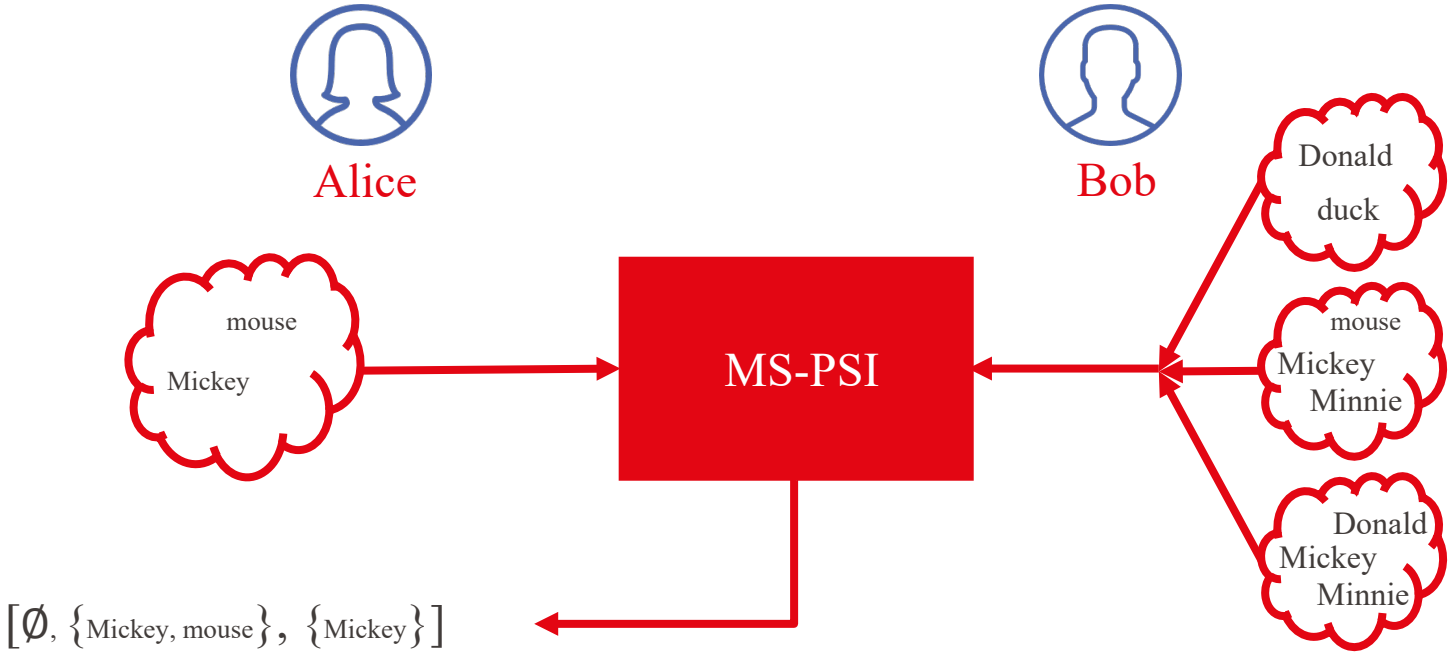
- The query
- The corpus
- The result
- The identities

Existing

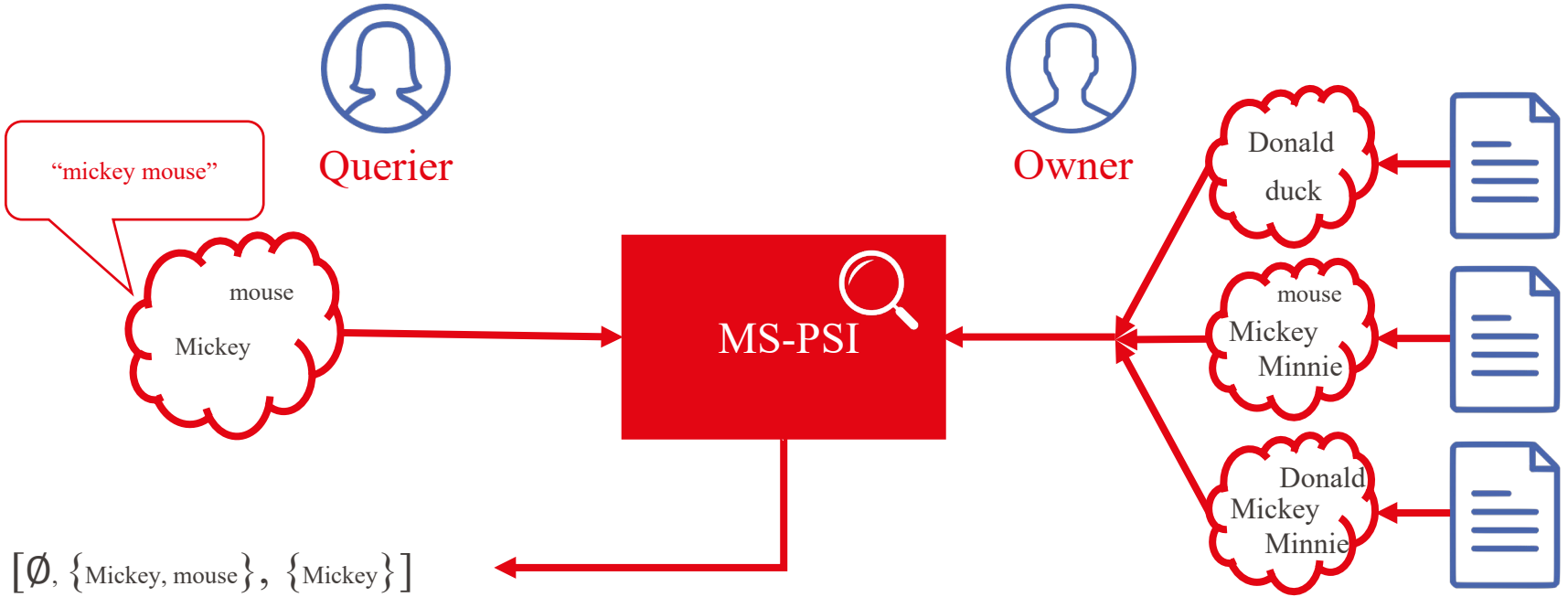
Search: Private Set Intersection (PSI)



Search: Multiset Private Set Intersection (MS-PSI)



Search: Multiset Private Set Intersection (MS-PSI)

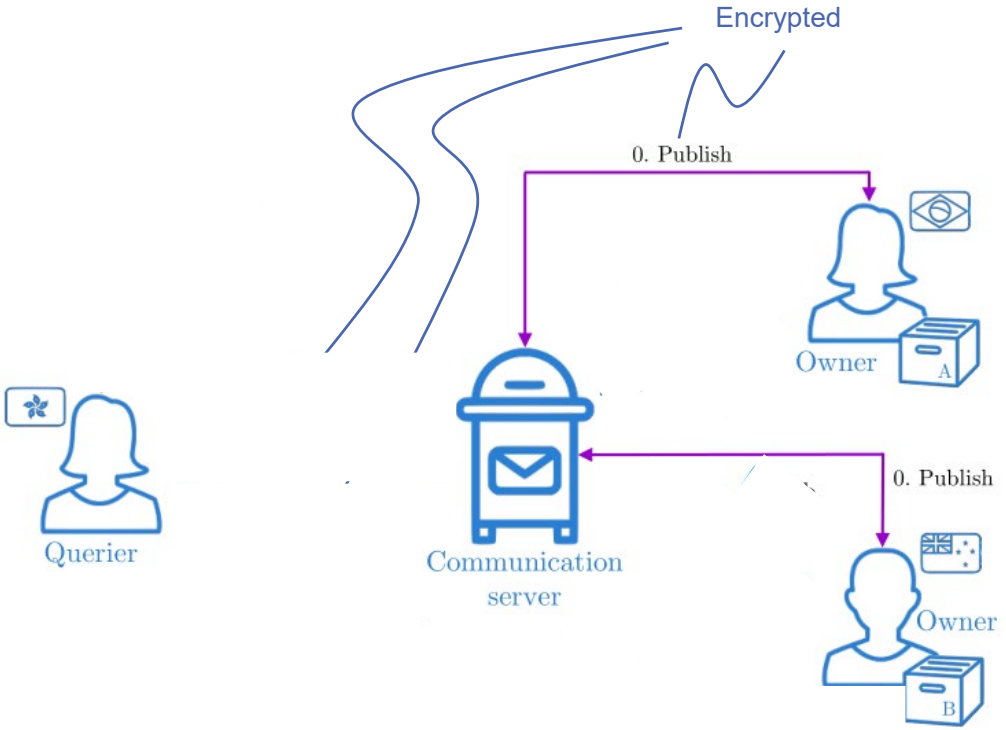


How search happens

Tokens permit others to verify that queries are from legitimate users (ICIJ associates).

Queries are sets of Named Entities. Their content **is secret** (collection owners do not know what has been queried).

Journalists querying the system learn if there are collections that match their query, but not from whom.
If collections do not match, journalists **do not learn anything**.

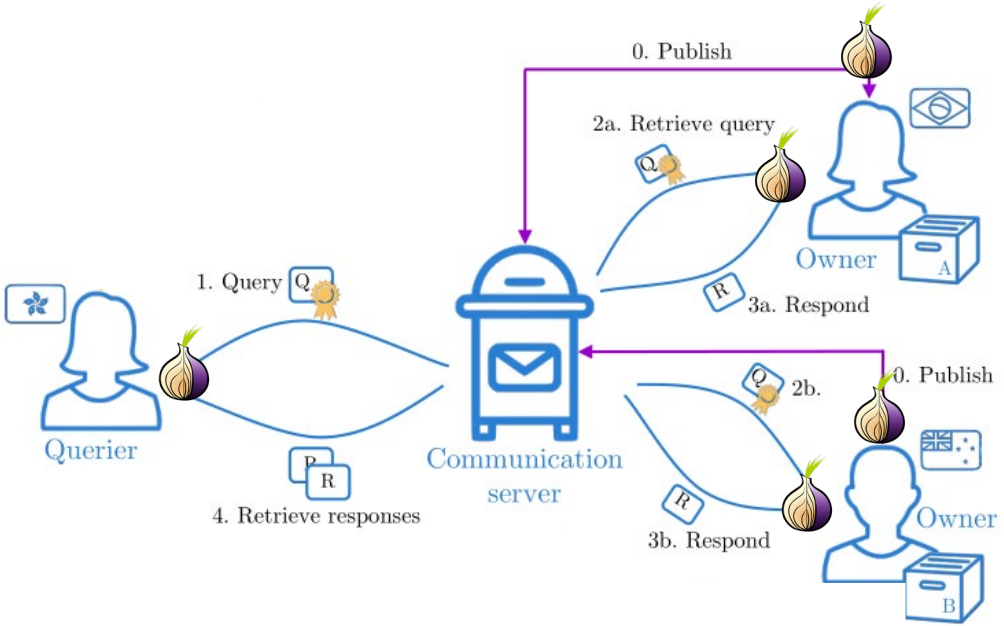


Protecting identities

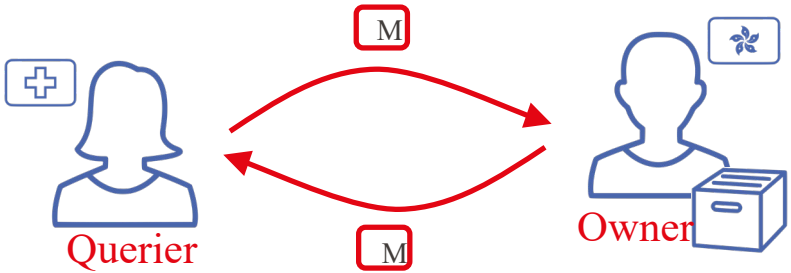
Journalists' identities are protected by the tokens and by the use of anonymous communications.

Neither ICIJ nor other journalists know the origin of the queries

Use of Tor a well-known anonymity network, to hide the IP addresses of the users



Private conversation for screening



Protect:
- The identities
- The existence of the conversation

Existing solutions can't hide the existence of the conversation

Existing infrastructure comes with constraints! (sounds familiar?)

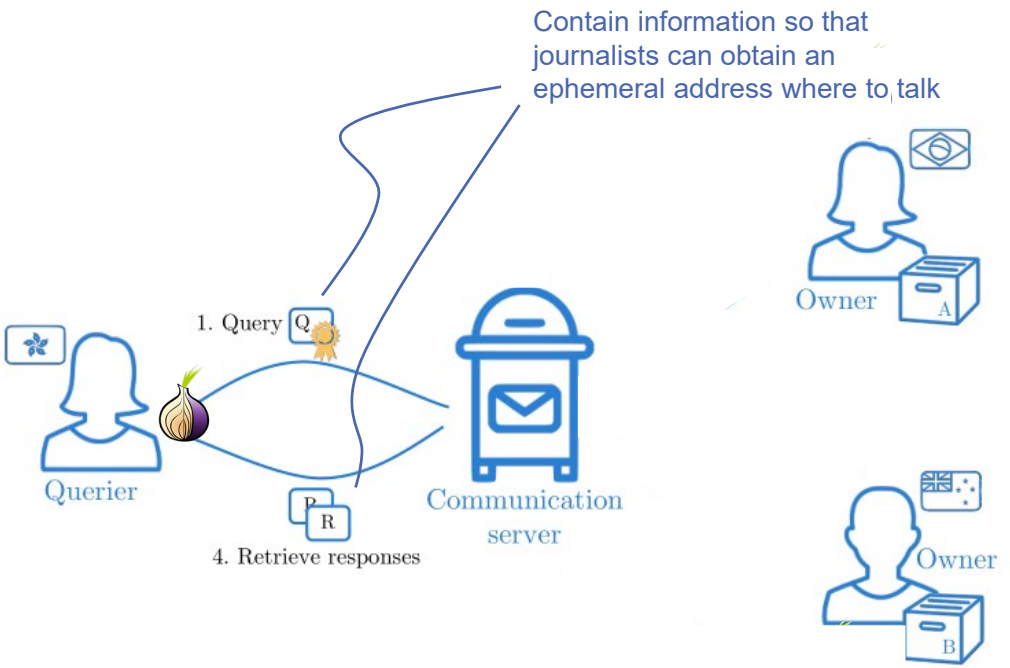
Anonymous conversations

After a match, journalists want to get access to the documents. Journalists may want to screen with whom they share documents.

Datashare Network enables **anonymous conversations** for screening.

Datashare Network **hides whether journalists are in a conversation**, to hide whether matches were found.

This limits the number of conversation messages per day.

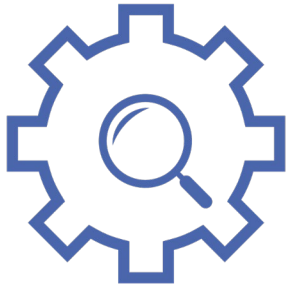


Privacy: layers

MS-PSI

Document search

Messaging



DatashareNetwork

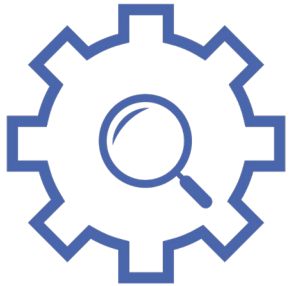
Privacy: layers

MS-PSI

Same privacy as PSI*

Document search

Messaging



DatashareNetwork

Privacy: layers

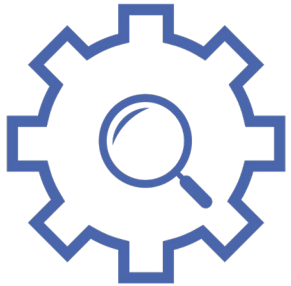
MS-PSI

Same privacy as PSI*

Document search

Only searched keywords

Messaging



DatashareNetwork

Privacy: layers

MS-PSI

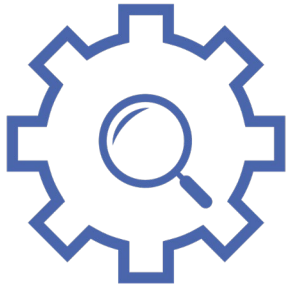
Same privacy as PSI*

Document search

Only searched keywords

Messaging

Unobservable



DatashareNetwork

Privacy: layers

MS-PSI

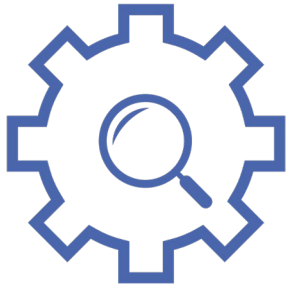
Same privacy as PSI*

Document search

Only searched keywords

Messaging

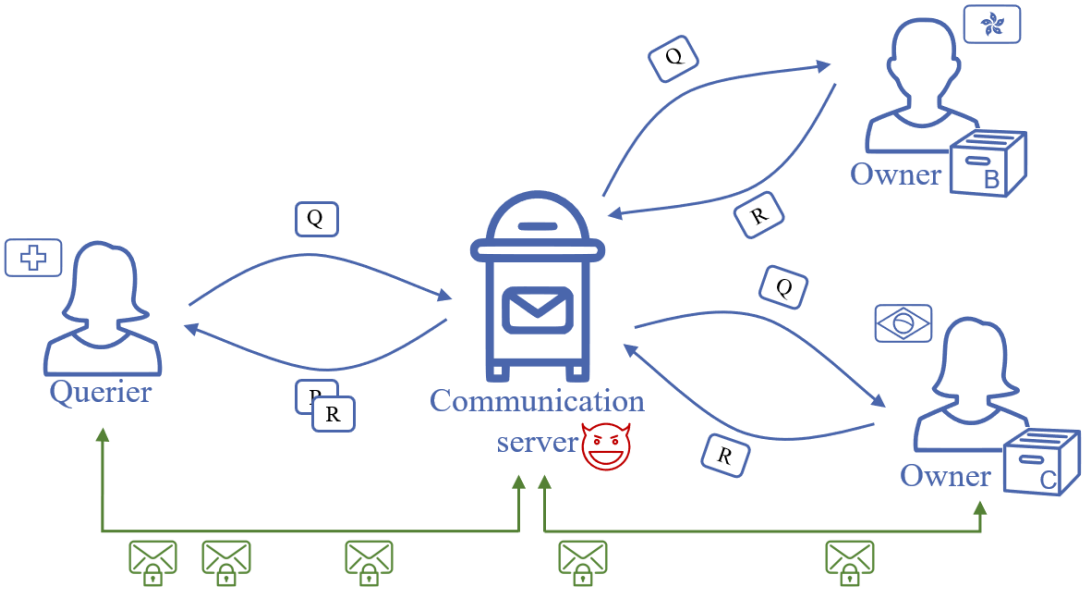
Unobservable



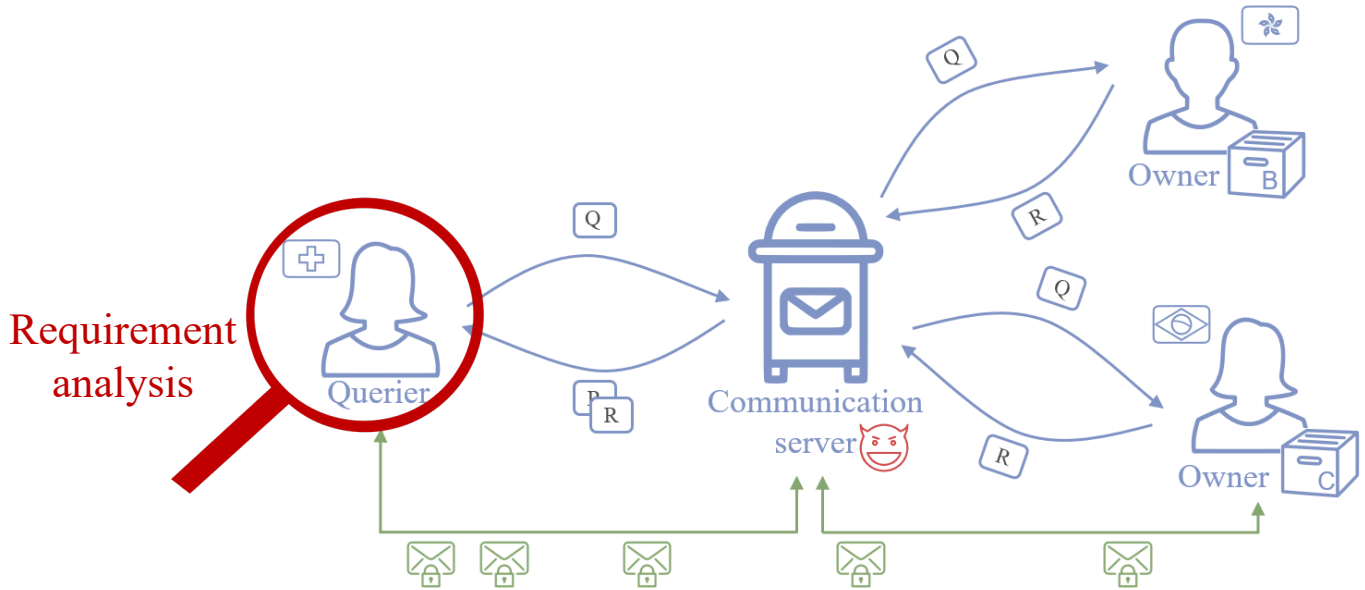
DatashareNetwork

Achieves privacy goals

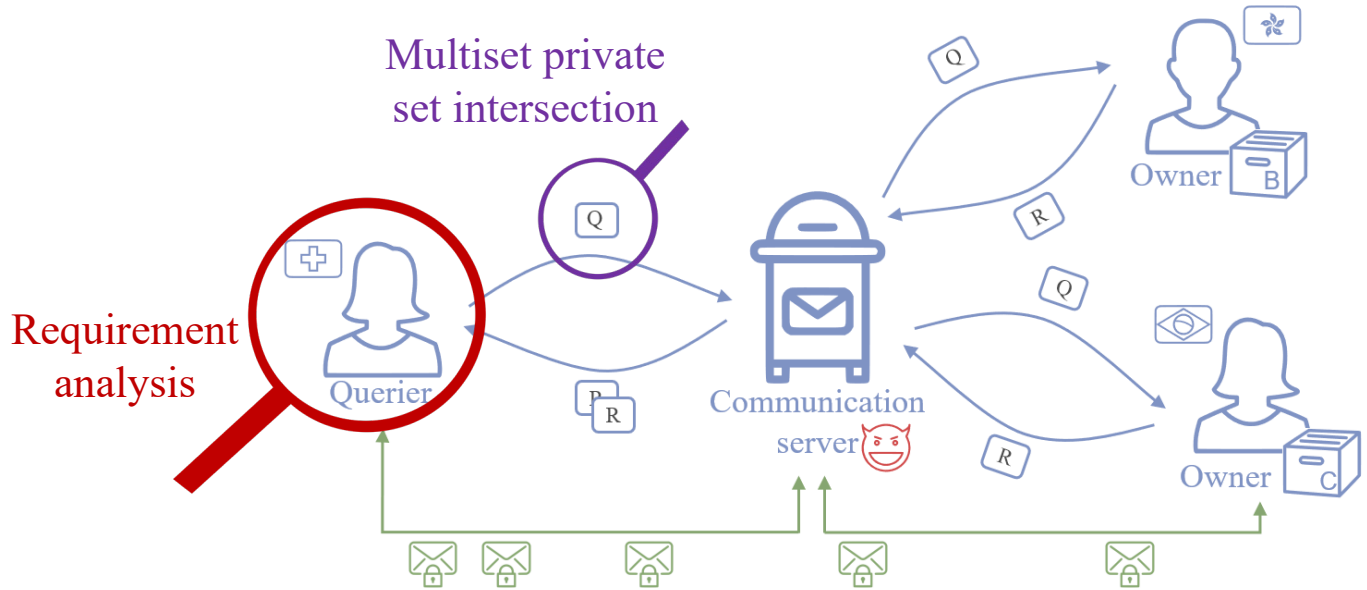
DatashareNetwork



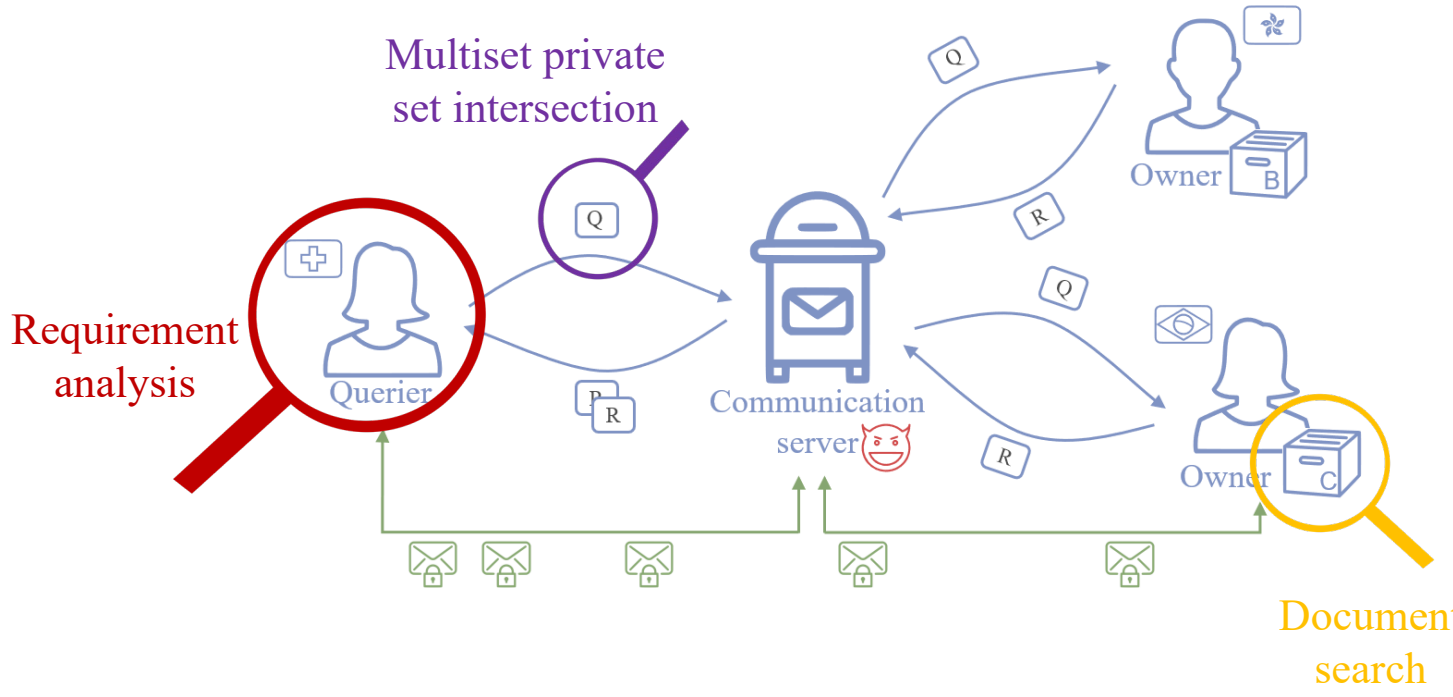
DatashareNetwork



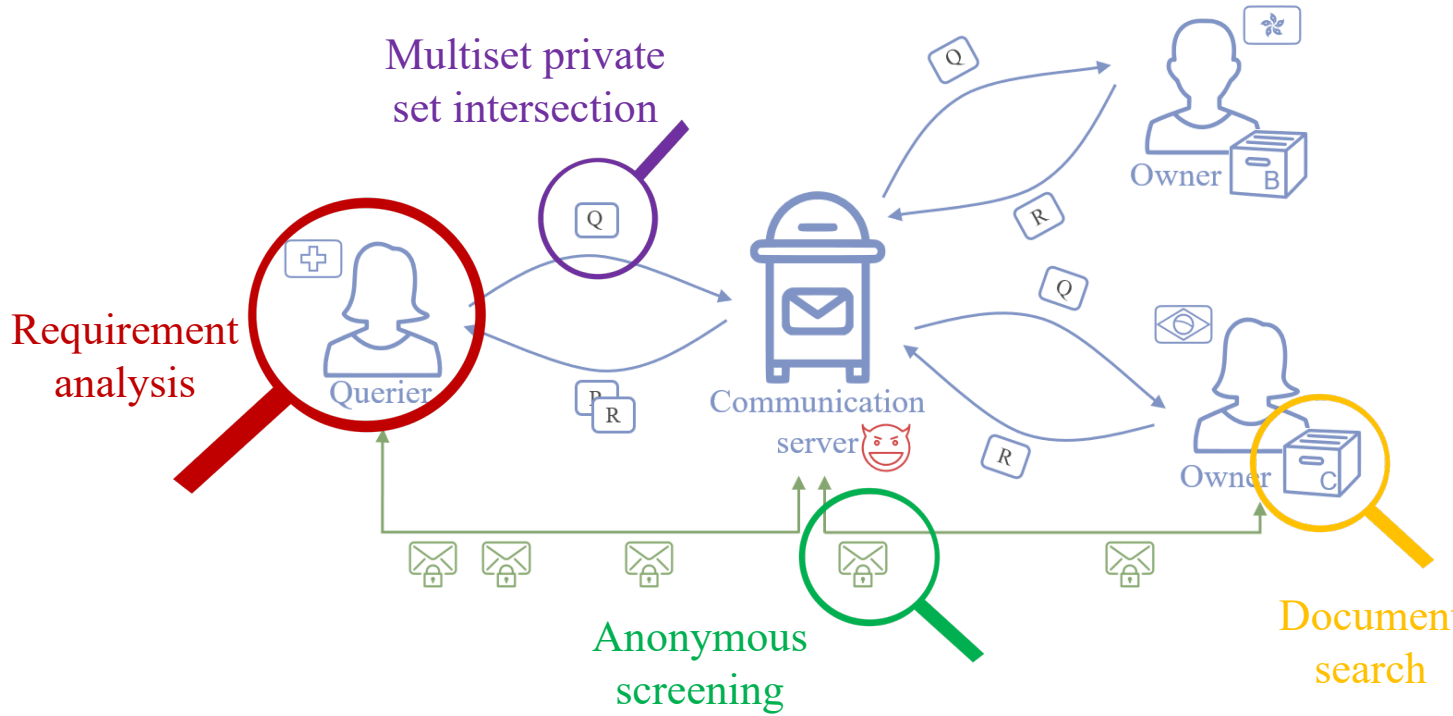
DatashareNetwork



DatashareNetwork



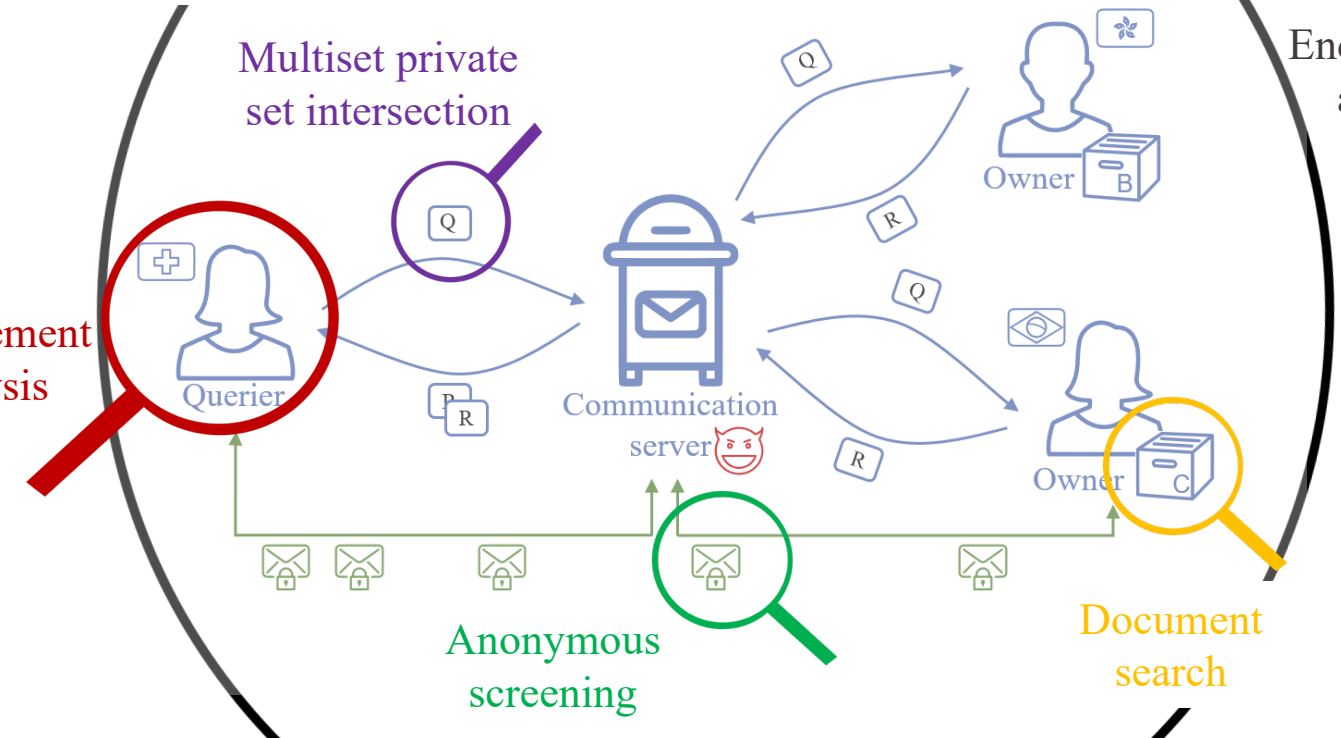
DatashareNetwork



ICIJ

EPFL

DatashareNetwork

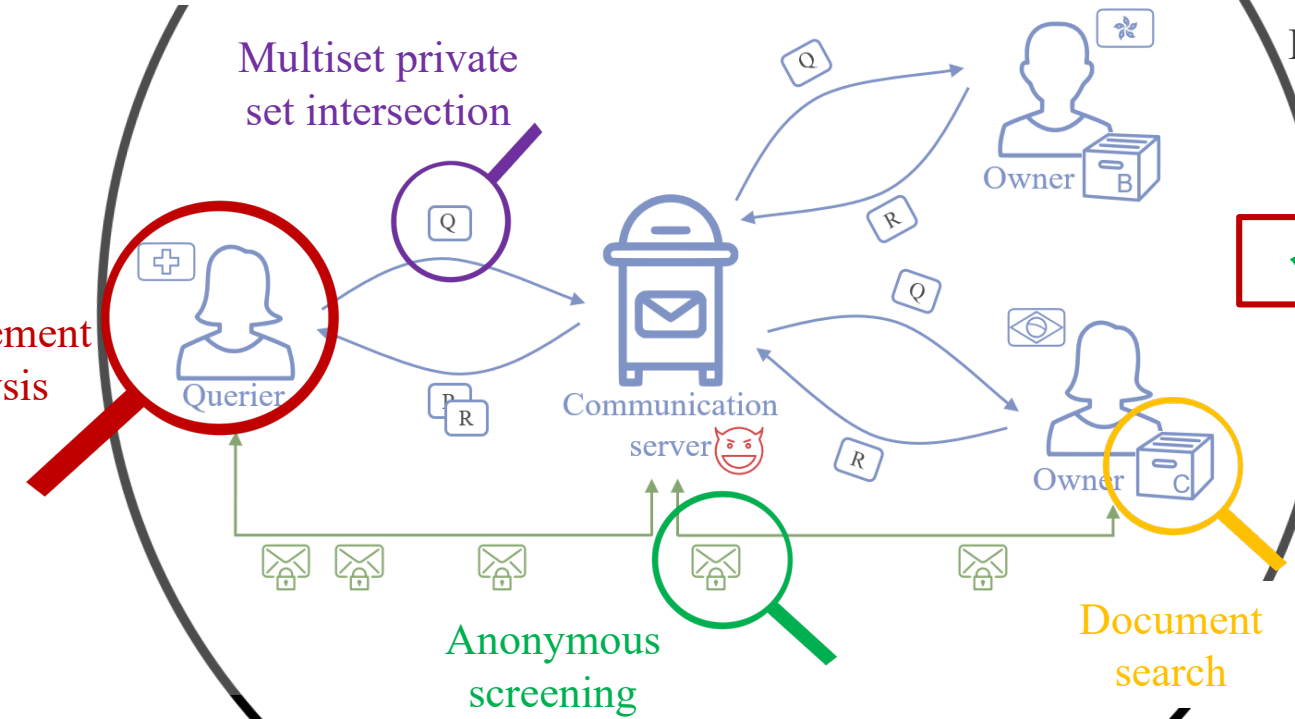


End-to-end system design
and privacy analysis

ICIJ

EPFL

DatashareNetwork



End-to-end system design and privacy analysis

✓ Formal proofs

ICIJ

EPFL

Summary

- **Fully-fledged decentralized search engine** inspired by stakeholders needs

- Key lessons
  - Requirement gathering is an iterative process
  - NGOs have different requirements than big companies: plenty of space for privacy technologies

- Steps ahead
  - User study
  - Deployment!

Thank you for your attention

Carmela Troncoso

École
polytechnique
fédérale
de Lausanne