CS-523 Advanced topics on Privacy Enhancing Technologies

Privacy engineering

Theresa Stadler
SPRING Lab
theresa.stadler@epfl.ch

Slides credit: Carmela Troncoso, Seda Guerses

Introduction
Privacy engineering

Course aim: learn **toolbox for privacy engineering**

You have a toolbox
How to use it?

Application Layer

Network Layer

Goals

What should you learn today?

- Understand the principles that guide **privacy-preserving design**

- Understand that privacy technologies alone are often **not enough to avoid all harms**

- Understand what makes privacy engineering **hard** in the real world

The goal: Privacy by design

“**Privacy by design is embedded into the design** and architecture of IT systems [...]. It is **not bolted as an addon**, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system without diminishing functionality".

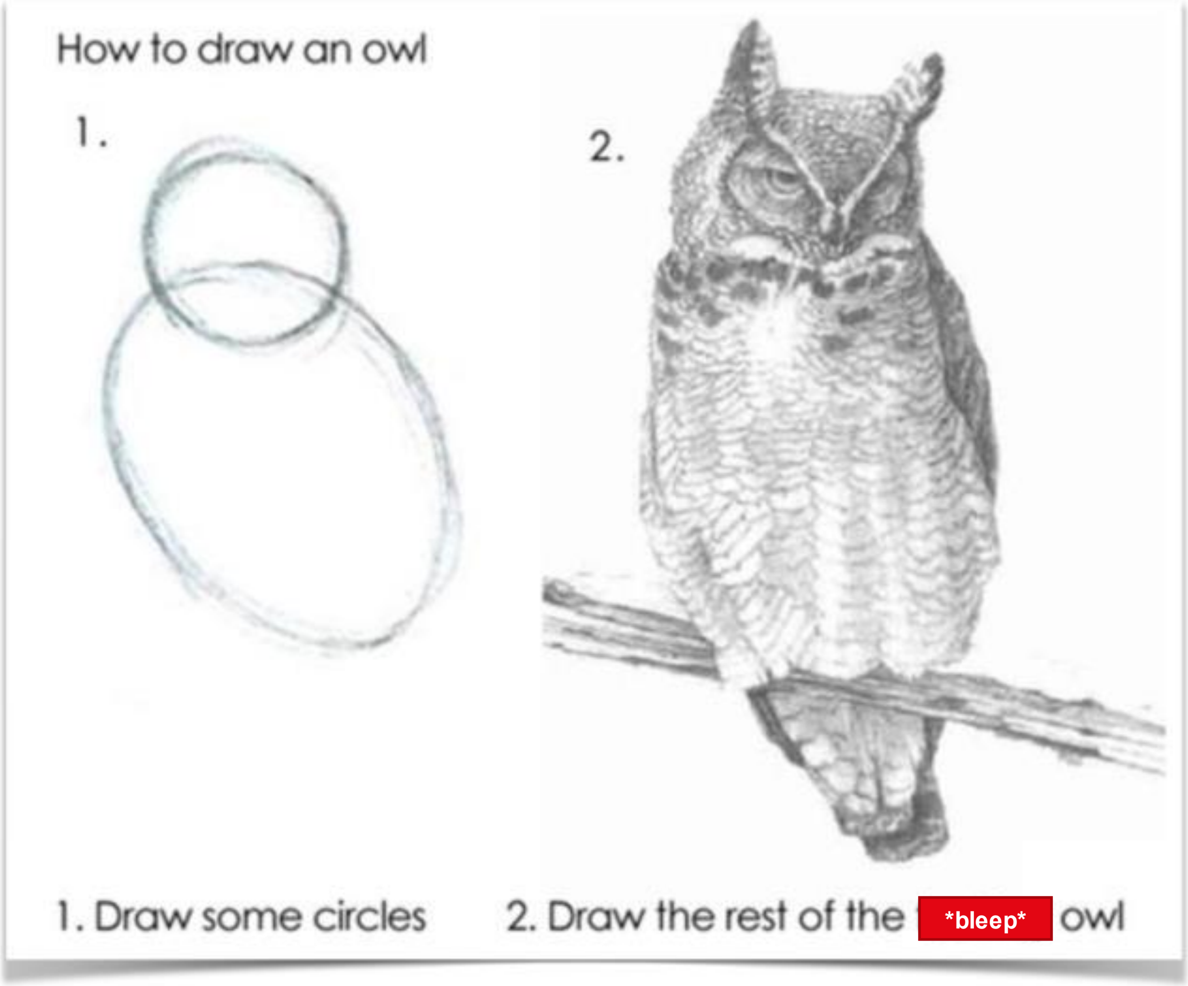
“the controller shall [...] **implement appropriate technical and organisational measures […] which are designed to implement data-protection principles** [...] in order to meet the requirements of this Regulation and protect the rights of data subjects.”

Companies should promote consumer privacy throughout their organizations and at every stage of the development of their products and services. Companies **should incorporate substantive privacy protections into their practices, such as data security, reasonable collection limits, sound retention practices, and data accuracy**.

The goal: Privacy by design

Privacy by Design

“**Privacy by design is embedded into the design and architecture of IT systems**
[...]. It is **no**
an essentia
to the syste



ational
nciples [...]
s of data

GDPR
General Data Protection Regul

Companies
stage of the
substantiv
reasonabl

every
orate

Privacy as data minimization

Build systems without data!
The least data in the system, the more privacy-preserving it is

→ Clearly related to a regulation principle

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

Privacy as data minimization

Build systems without data!
The least data in the system, the more privacy-preserving it is

→ Clearly related to a regulation principle

But, **it's not "data" that is minimized** (in the system as a whole)

Data is kept on user devices
Data is sent encrypted to a server (only client has the key)
Data is distributed over multiple servers
…

"data minimization" alone is a BAD metaphor for privacy-preserving designs

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

Privacy as trust minimization

Build systems that minimize **privacy risks and trust assumptions** placed on other entities

→ Limit unintended uses of data by untrusted entities

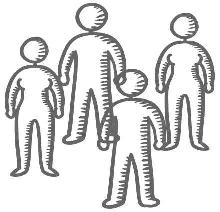
Seda Gurses, Carmela Troncoso, Claudia Diaz. Engineering Privacy by Design.Computers, Privacy & Data Protection. 2011

Privacy as trust minimization

Build systems that minimize **privacy risks and trust assumptions** placed on other entities

→ Limit unintended uses of data by untrusted entities

Who are these ”untrusted entities”?



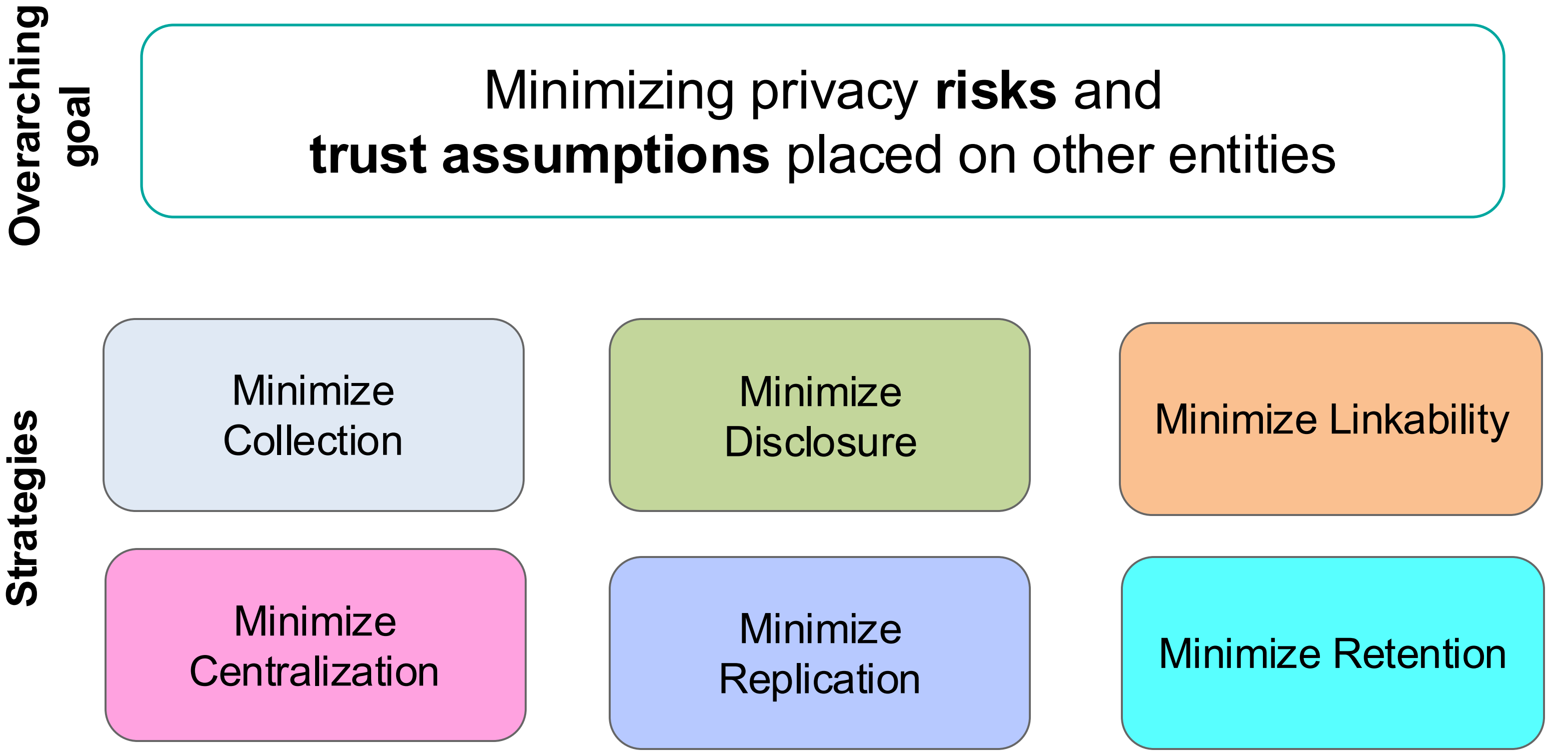
other users
third parties

semi-trusted
service provider

malicious
service provider

Seda Gurses, Carmela Troncoso, Claudia Diaz. Engineering Privacy by Design Reloaded. Amsterdam Privacy Conference. 2015
Seda Gurses and Claudia Diaz. "Two tales of privacy in online social networks." IEEE Security & Privacy Magazine. 2013

Privacy as trust minimization



Overarching goal

Minimizing privacy **risks** and
trust assumptions placed on other entities

Strategies

Minimize Collection

Minimize Disclosure

Minimize Linkability

Minimize Centralization

Minimize Replication

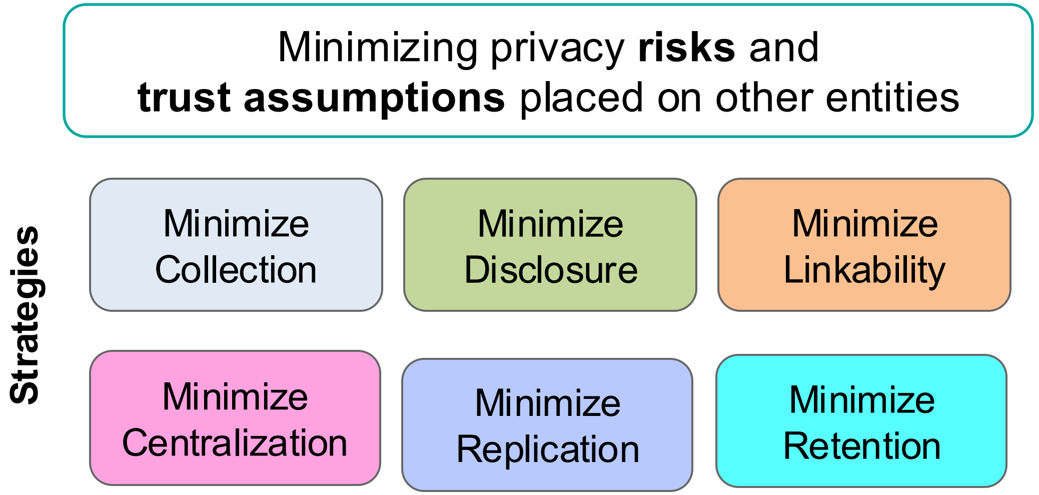
Minimize Retention

For data **and** metadata

- Seda Gurses, Carmela Troncoso, Claudia Diaz. Engineering Privacy by Design Reloaded. Amsterdam Privacy Conference. 2015

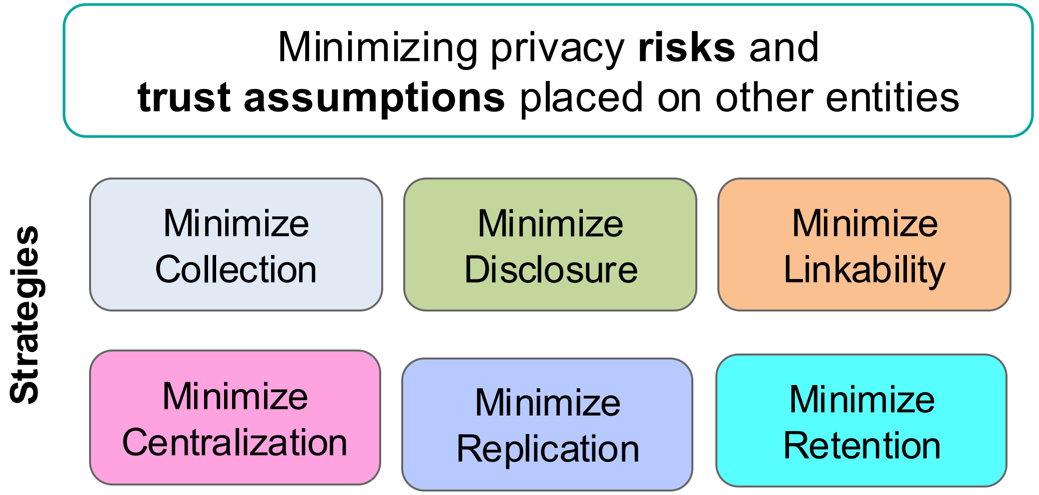
Technological solutions to implement these strategies

- do not send the data (local computations)
- encrypt the data
- use advanced privacy-preserving protocols
- obfuscate the data
- anonymize the data



Technological solutions to implement these strategies

- do not send the data (local computations)
- encrypt the data
- use advanced privacy-preserving protocols
- obfuscate the data
- anonymize the data



Minimizing privacy **risks** and **trust assumptions** placed on other entities

Strategies

Minimize Collection

Minimize Disclosure

Minimize Linkability

Minimize Centralization

Minimize Replication

Minimize Retention

But minimizing trust does not guarantee that we minimize harm.

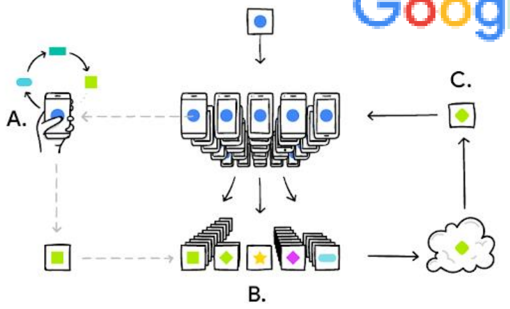
What if the purpose(s) of the application is harmful?

Technological solutions to implement these strategies

Example: Privacy-preserving online content personalisation

- Trend towards On-Device learning and encryption in targeted advertising and content personalisation
- Promoted as “privacy-by-design” approaches
- Harms of targeting and personalisation (filter bubbles, discrimination,…) persist

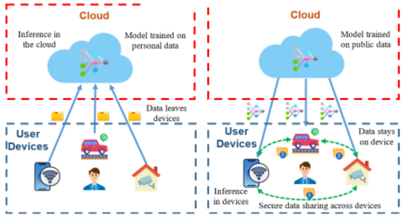
Google Research



Privacy-Enhancing Technologies and Building for the Future

Meta

Samsung Research



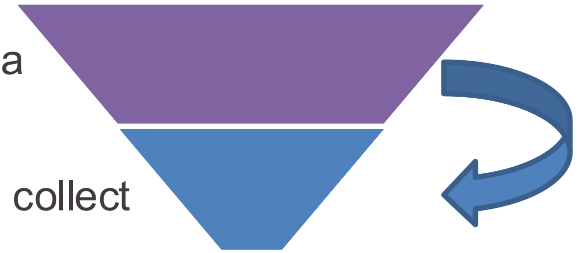
But minimizing trust does not guarantee that we minimize harm.

What if the purpose(s) of the application is harmful?

- https://about.fb.com/news/2020/10/a-path-forward-for-privacy-and-online-advertising/
https://research.samsung.com/blog/Resolving-Privacy-Personalization-Paradox

Privacy by design thinking process

The usual approach in the past

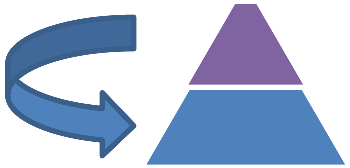


I want all data

Data I can collect

Data protection compliance

The privacy engineering approach



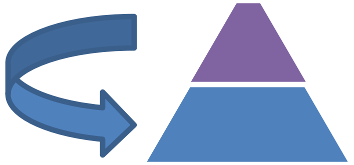
Operational purposes

Data needed for the **purpose**

Data I will finally collect

Privacy by design thinking process

The privacy engineering approach



Operational purposes

PETs

Data needed for the **purpose**

Data I will finally collect

Build systems that limit data use to the **intended purpose**

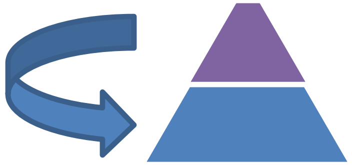
Purpose limitation is a **good** metaphor for privacy-preserving designs

The hard bit... What is the purpose?

The privacy engineering approach

Operational purposes

PETs



Data needed for the **purpose**

Data I will finally collect

? ? ? ?

What do you need?

Everything?!

COMITE INTERNATIONAL GENEVE
ICRC

BAG
OFSP
UFSP

Purpose limitation is a **good** metaphor for privacy-preserving designs

The hard bit… What is the purpose?

Help stakeholders understand what the purpose of the system is

What do you **do**?

…

And **why** do you do this?

…

→ Purpose of the system may be **broad** or actually comprise multiple purposes

The hard bit… What is the purpose?

→ Purpose of the system may be **broad** or comprise **multiple** purposes

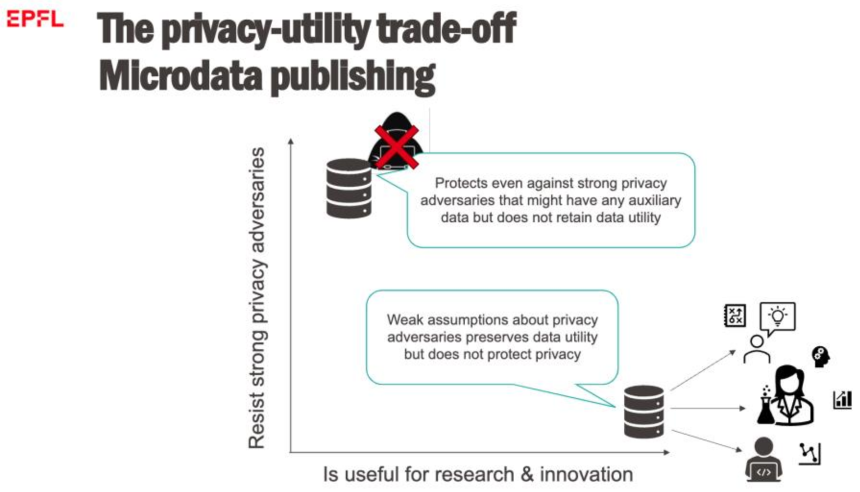
Purpose limitation becomes really hard!

- ➢ Combination of inputs per purpose may enable more uses/purposes than intended

Your role as privacy engineer:

- ✓ Quantify potential harms (to the extent possible)
- ✓ Explain the risks to the stakeholder

Week 6: Data publishing



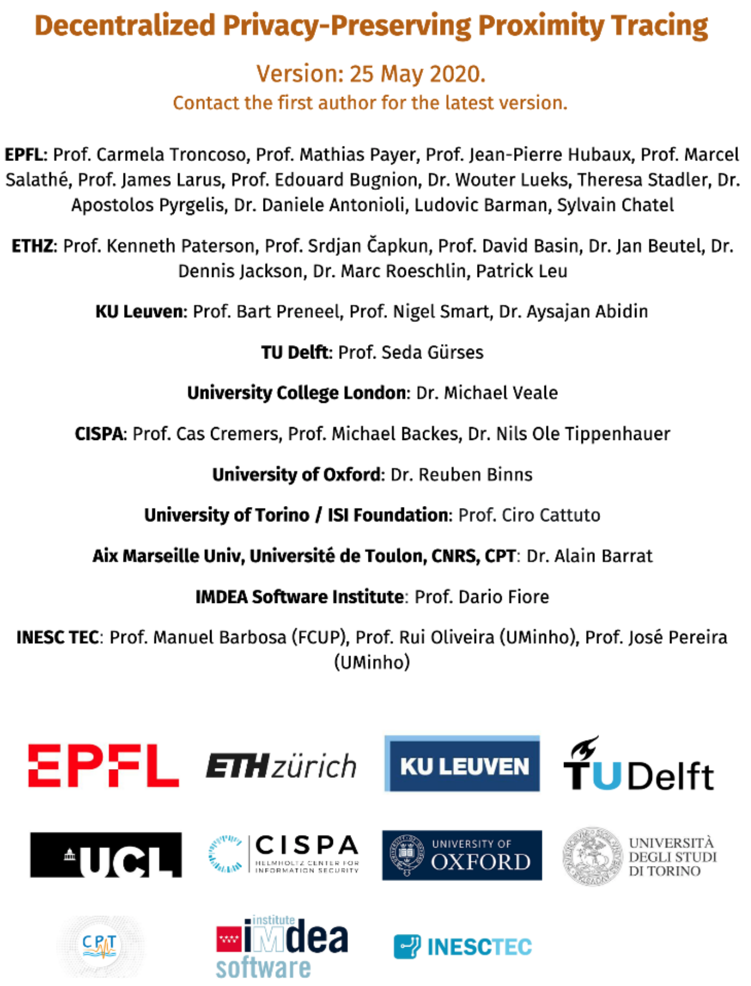
The privacy-utility trade-off
Microdata publishing

Protects even against strong privacy adversaries that might have any auxiliary data but does not retain data utility

Weak assumptions about privacy adversaries preserves data utility but does not protect privacy

EPFL
Resist strong privacy adversaries
Is useful for research & innovation

From the lab to deployment
Example: Digital proximity tracing



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

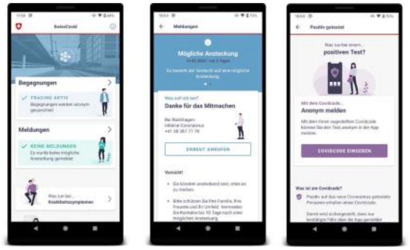
IMDEA Software Institute: Prof. Dario Fiore

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

EPFL ETHzürich KU LEUVEN TUDelft
UCL CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY UNIVERSITY OF OXFORD UNIVERSITÀ DEGLI STUDI DI TORINO
CPT institute imdea software INESCTEC

 | Google

GAEN framework



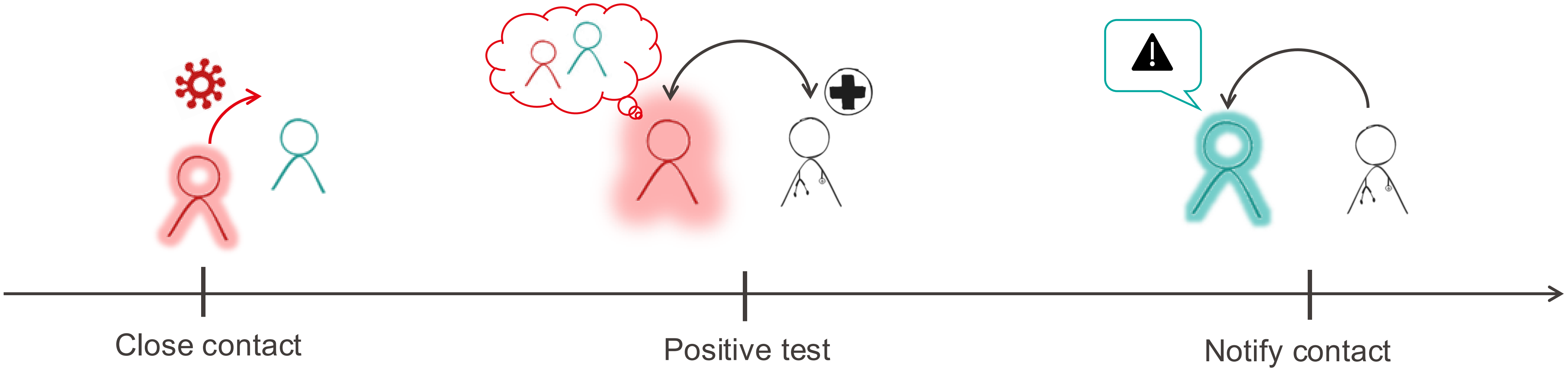
SwissCovid

43+ States / Countries

~100 million users

Contact Tracing

How to contain the spread of an infectious agent throughout a population?



Close contact

Positive test

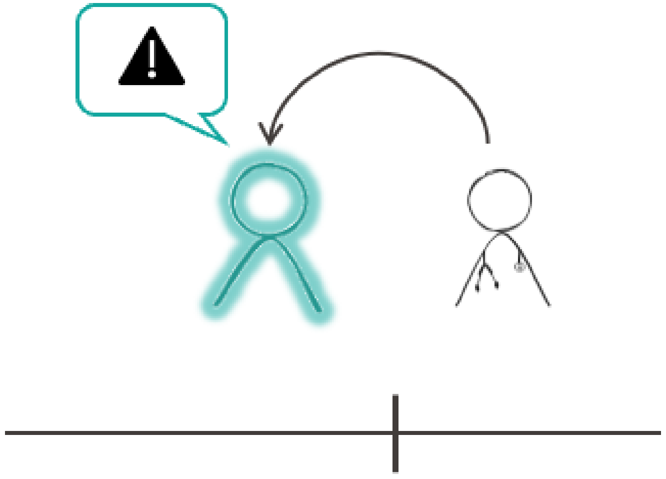
Notify contact

Notify at-risk contacts of past exposure to the infectious agent

Digital Proximity Tracing Systems

Purpose

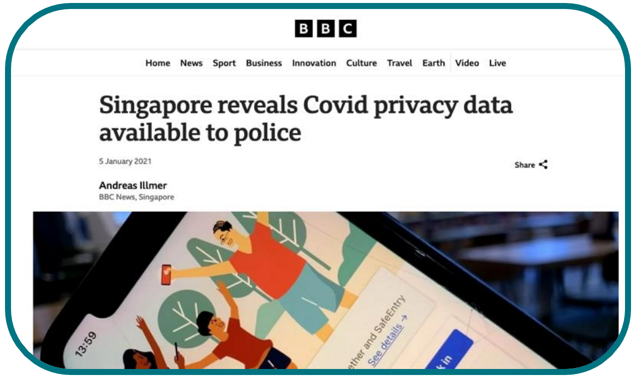
Notify at-risk contacts of past exposure to the infectious agent



Notify contact

And **NOT**

- Collect data on who interacted with whom
- Collect data on who went **where** and when
- Notify contacts who of their past interactions tested positive
- …


BBC
Home News Sport Business Innovation Culture Travel Earth Video Live
Share
13:59
ether and SafeEntry
See details →
k in

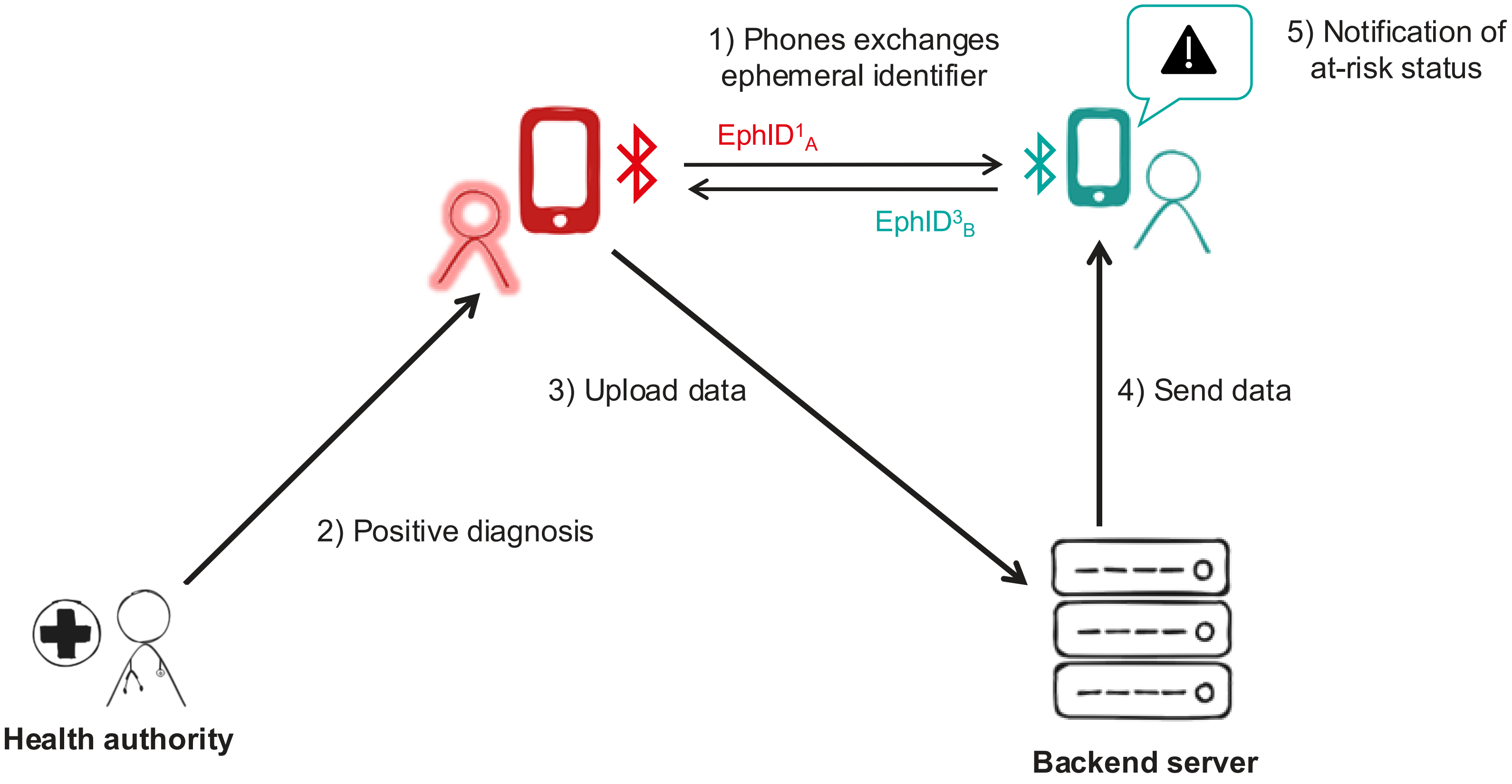
Singapore reveals Covid privacy data available to police

5 January 2021

Andreas Illmer
BBC News, Singapore

Harms of system misuse

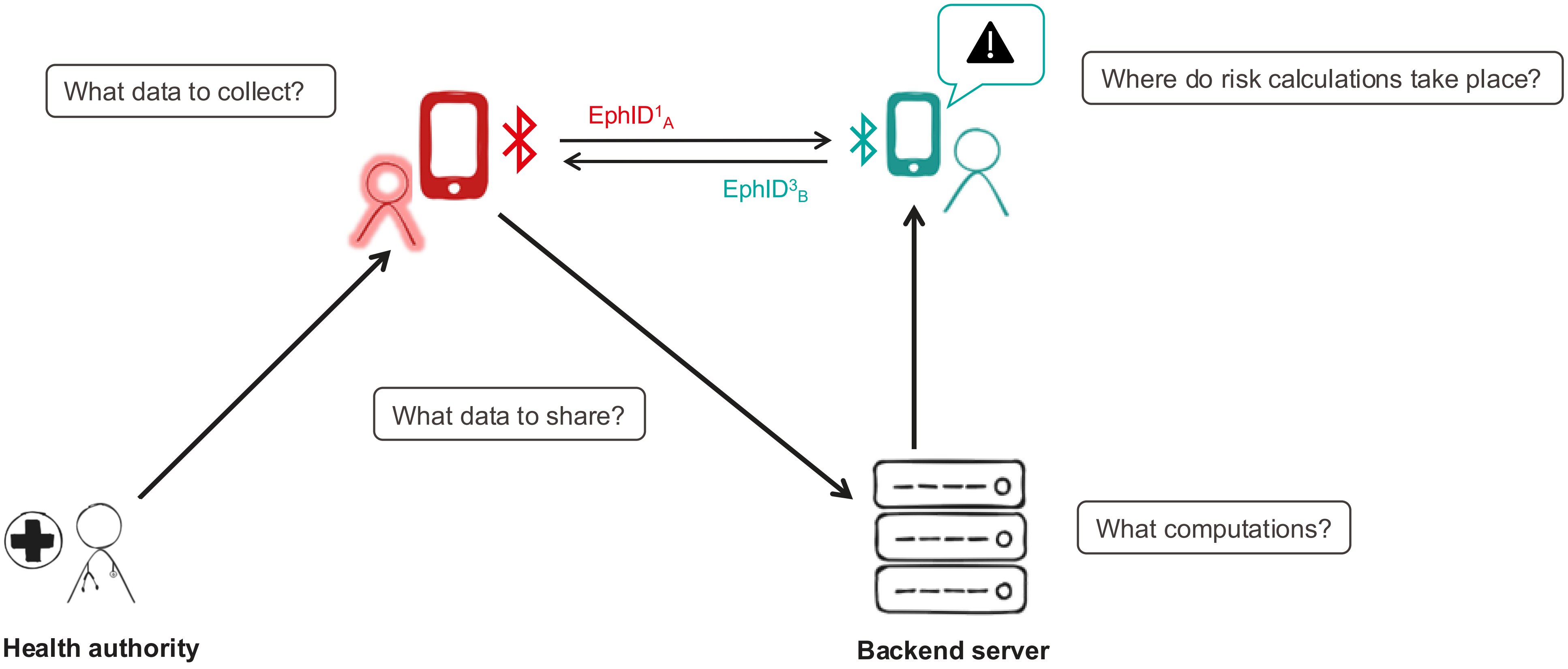
Digital Proximity Tracing Systems



Intended purpose: Provide a mechanism to alert at-risk contacts.

Digital Proximity Tracing Systems

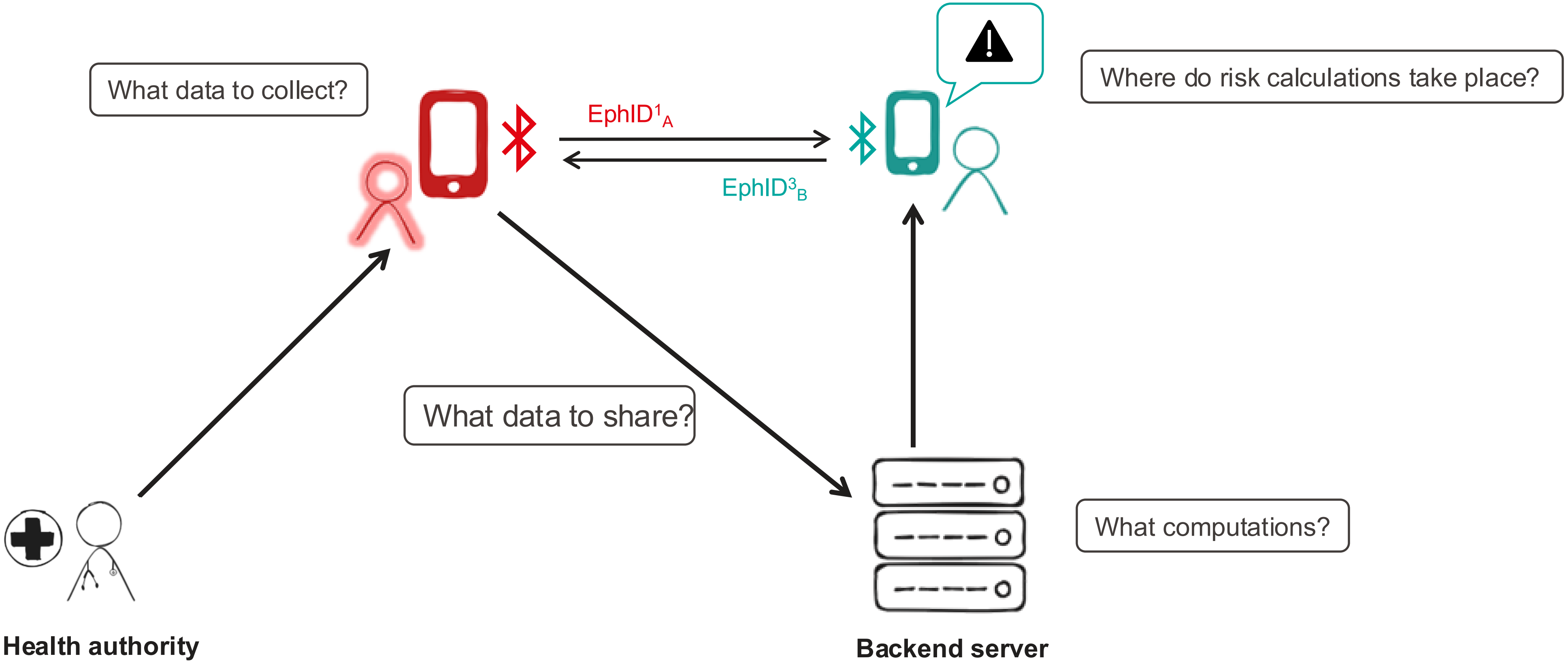
Design



Intended purpose: Provide a mechanism to alert at-risk contacts.

Inherent Risks

Core functionality: Notify at-risk contacts



What data to collect?

EphID^1_A

EphID^3_B

Where do risk calculations take place?

What data to share?

What computations?

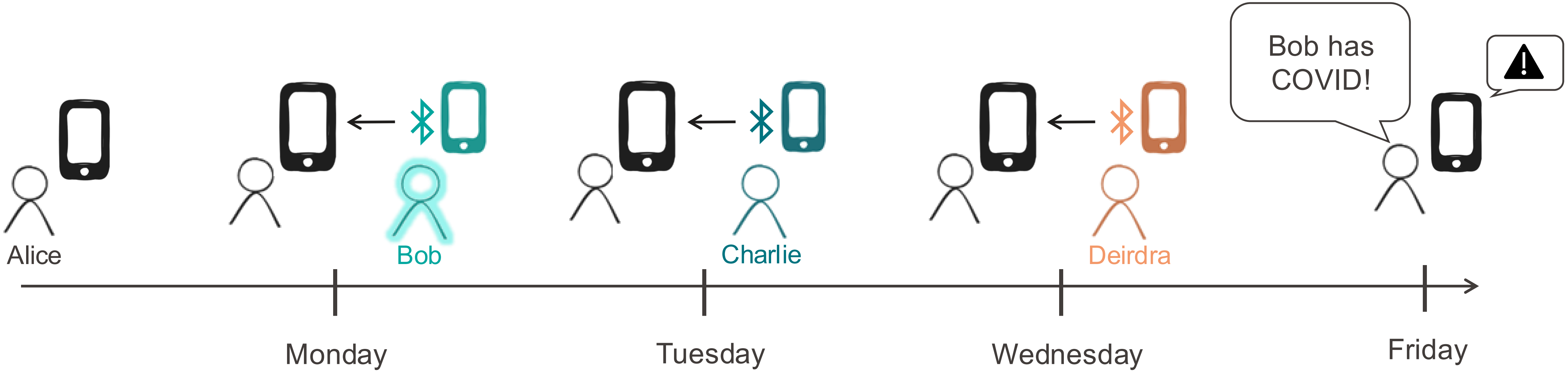
Health authority

Backend server

Intended purpose: Provide a mechanism to alert at-risk contacts.

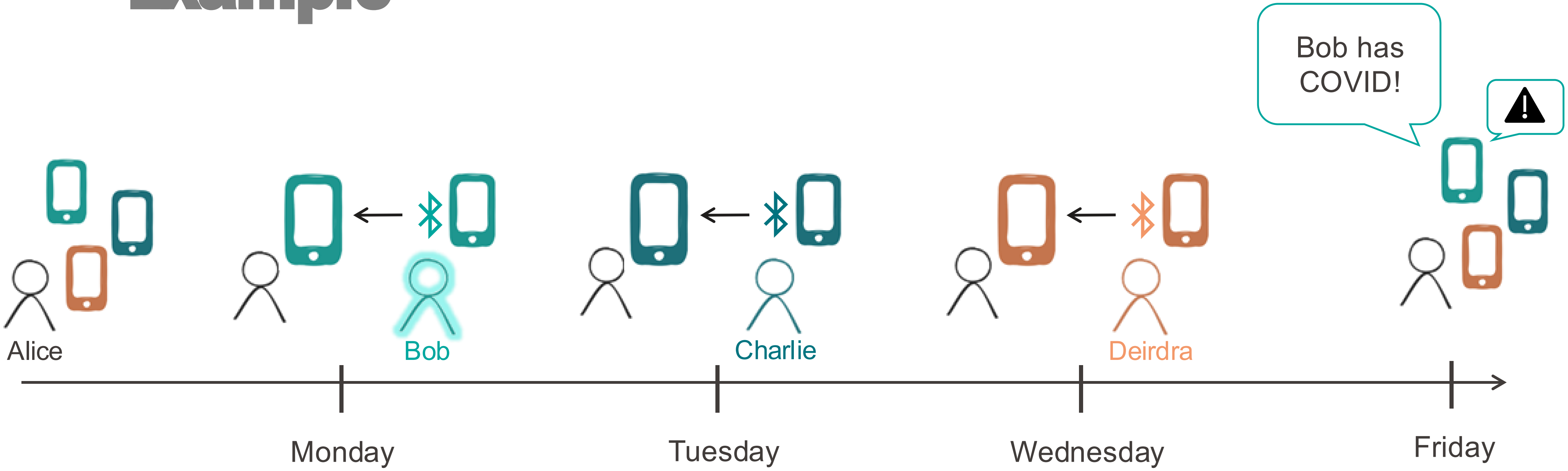
Inherent Risks

Example

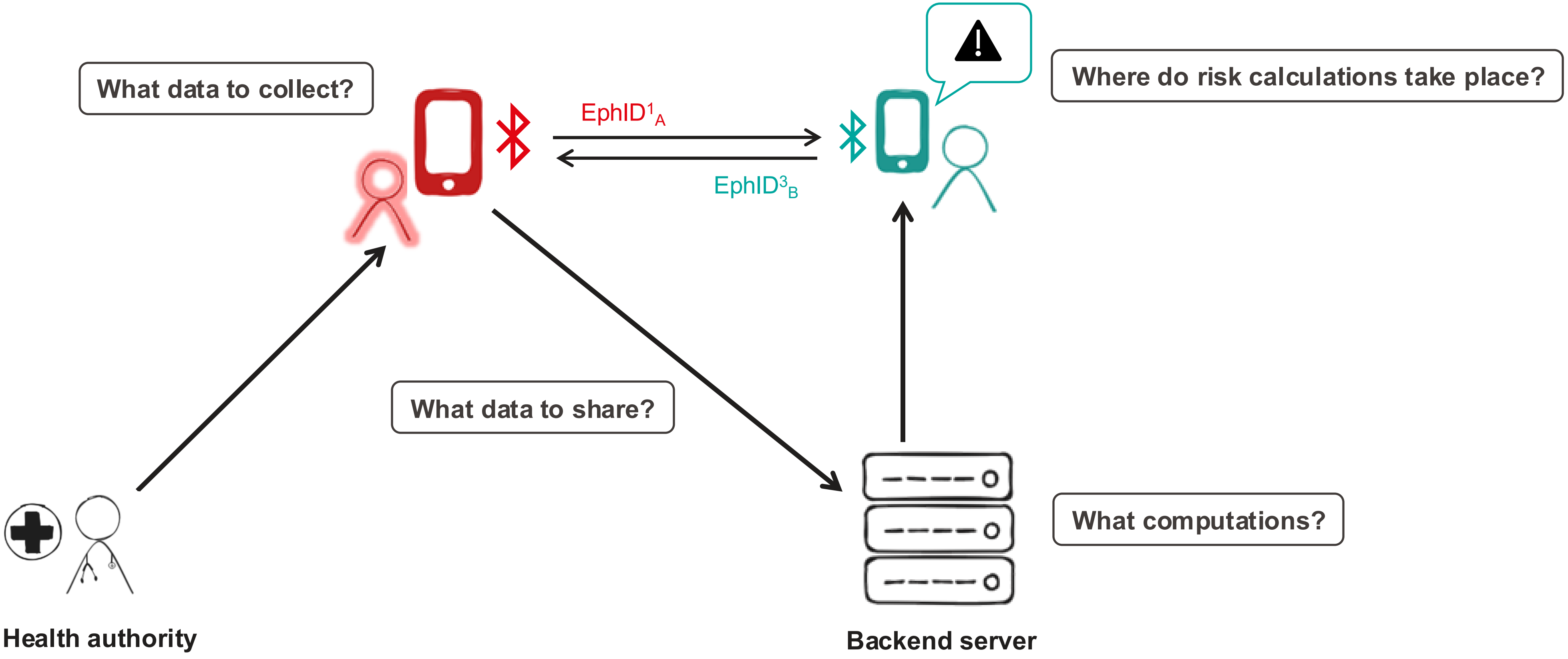


Inherent Risks

Example



Beyond Inherent Risks - Design Choices



What data to collect?

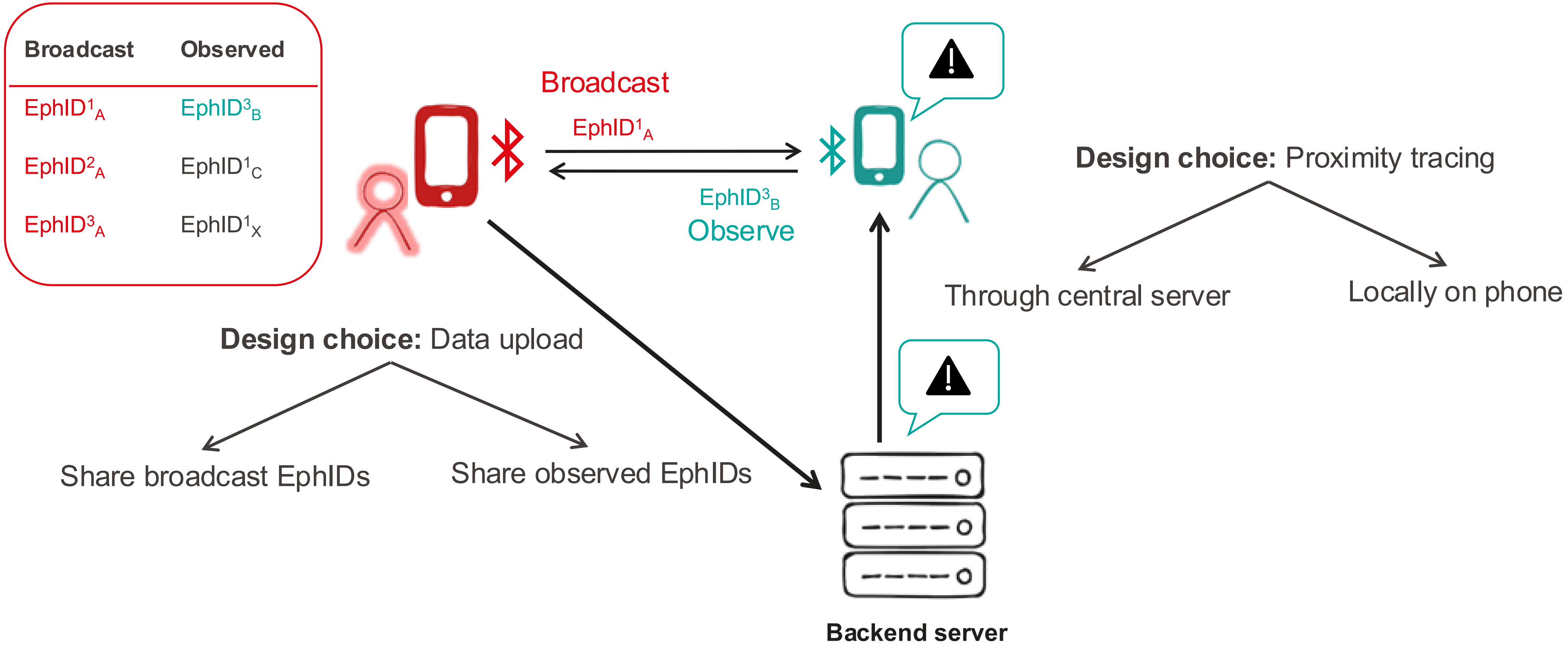
Where do risk calculations take place?

What data to share?

What computations?

Intended purpose: Provide a mechanism to alert at-risk contacts.

Beyond Inherent Risks - Design Choices



Broadcast	Observed
EphID^1_A	EphID^3_B
EphID^2_A	EphID^1_C
EphID^3_A	EphID^1_X

Design choice: Proximity tracing

- Through central server
- Locally on phone

Design choice: Data upload

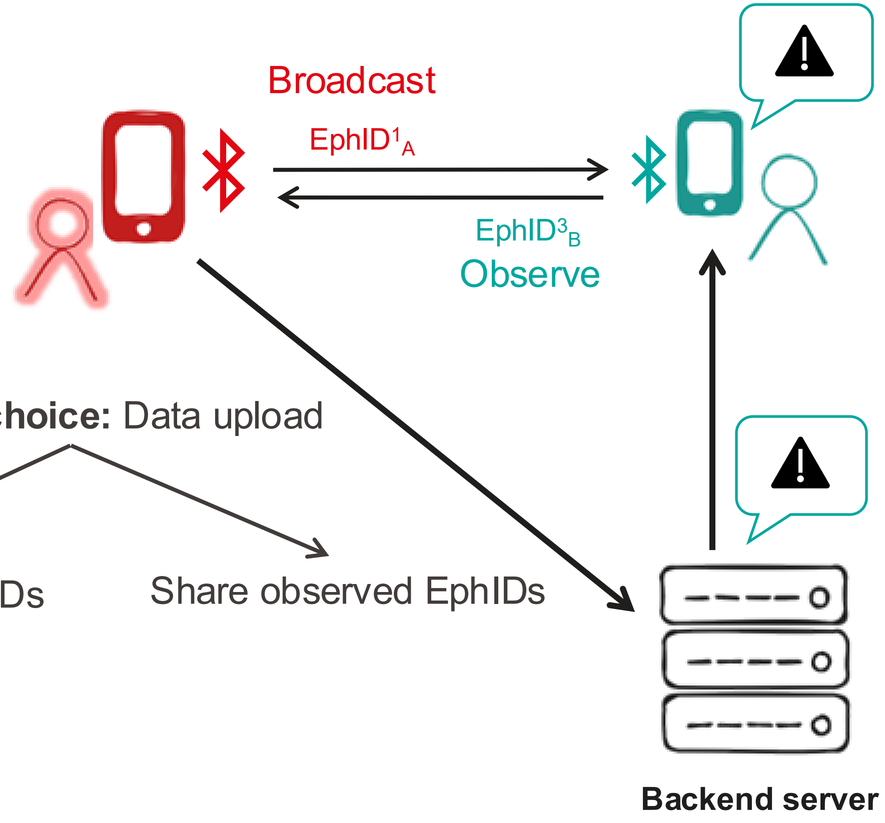
- Share broadcast EphIDs
- Share observed EphIDs

System Design

Given these choices, what would you do?

- What information can you extract from the observed/broadcasted ephemeral IDs?
- What is your threat model?
- What are the harms?
- How does your choice affect intended functionality?

Broadcast	Observed
EphID^1_A	EphID^3_B
EphID^2_A	EphID^1_C
EphID^3_A	EphID^1_X

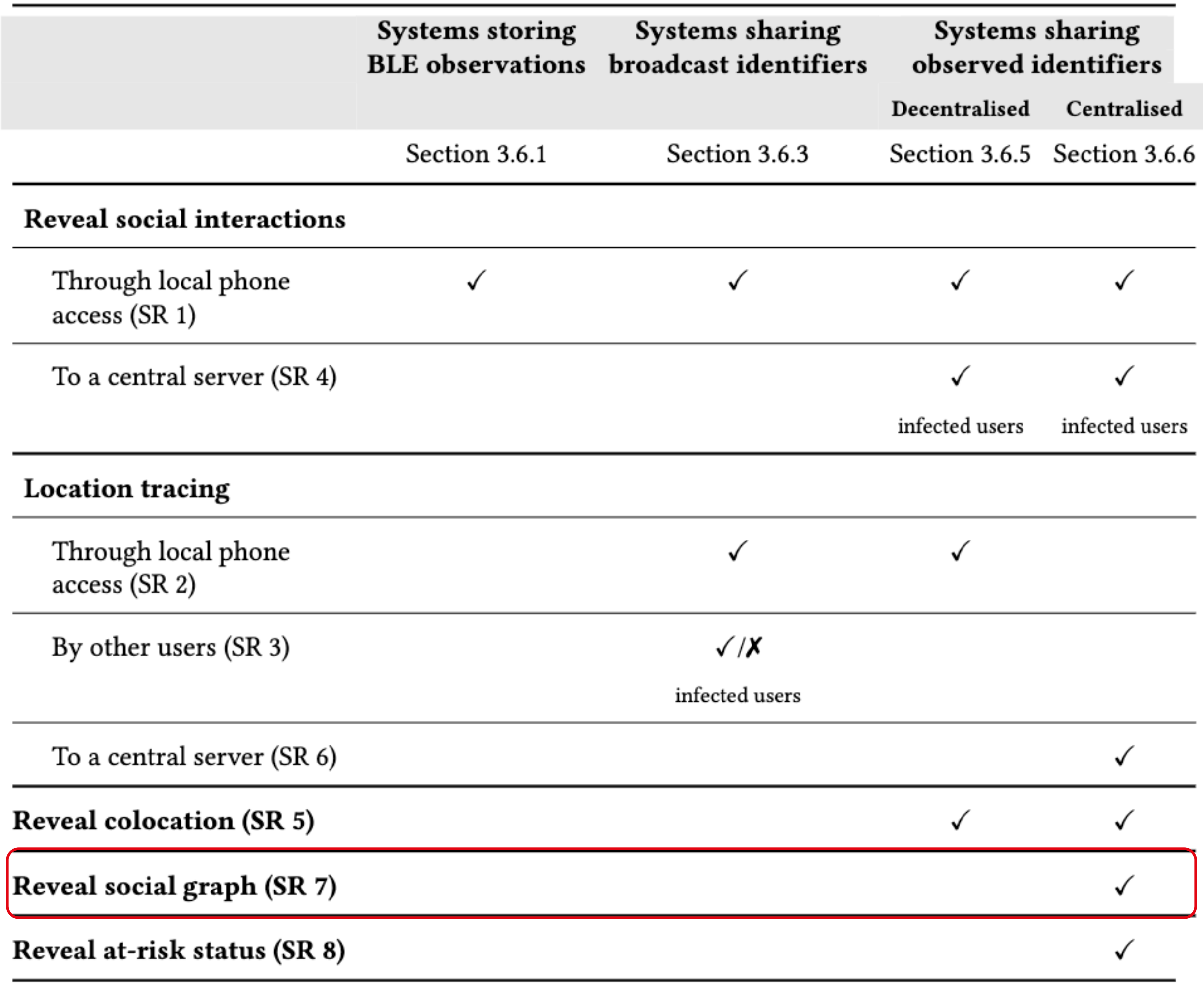


Design choice: Data upload

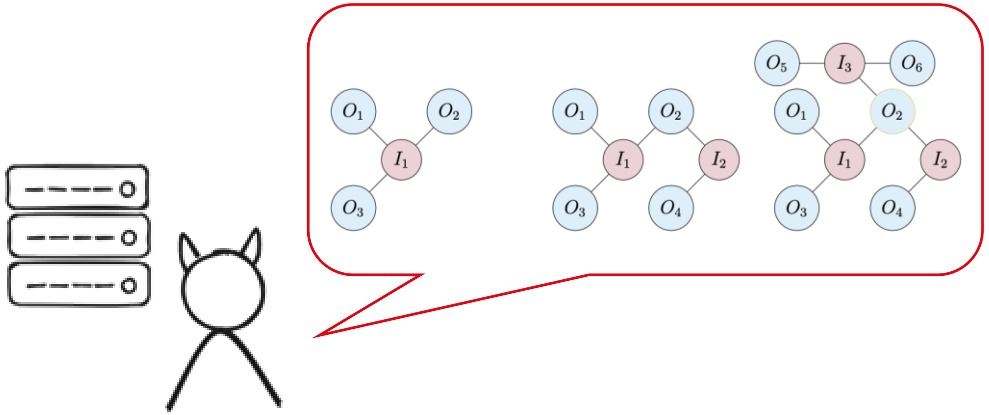
Share broadcast EphIDs

Share observed EphIDs

Beyond Inherent Risks - System Comparison



	Systems storing BLE observations	Systems sharing broadcast identifiers	Systems sharing observed identifiers	
			Decentralised	Centralised
	Section 3.6.1	Section 3.6.3	Section 3.6.5	Section 3.6.6
Reveal social interactions				
Through local phone access (SR 1)	✓	✓	✓	✓
To a central server (SR 4)			✓ infected users	✓ infected users
Location tracing				
Through local phone access (SR 2)		✓	✓	
By other users (SR 3)		✓/✗ infected users		
To a central server (SR 6)				✓
Reveal colocation (SR 5)			✓	✓
Reveal social graph (SR 7)				✓
Reveal at-risk status (SR 8)				✓



That was actually “just” the protocol…

SwissCovid has more privacy mechanisms that required a lot of engineering

- Privacy-preserving keys publication
- Dummy cover traffic to protect positive uploads
- Privacy-preserving statistics collection
- Privacy-preserving logging strategies
- …

Digital Proximity Tracing

Conclusions

- Even best privacy-preserving design cannot eliminate inherent risks linked to intended functionality

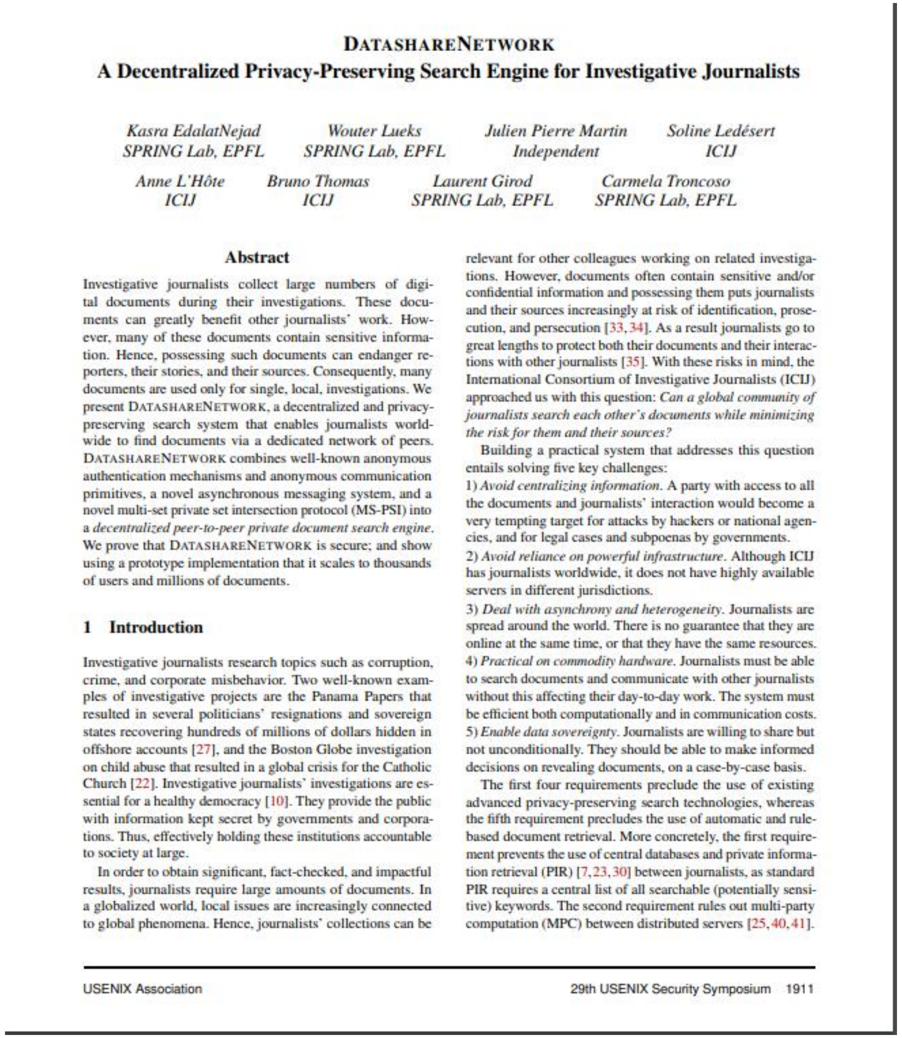
- Through risk analysis can identify major design decisions which facilitate system design

From the lab to deployment

Example: DataShare

SPRING

SECURITY AND PRIVACY ENGINEERING LABORATORY



DATASHARENETWORK
A Decentralized Privacy-Preserving Search Engine for Investigative Journalists

Kasra EdalatNejad
SPRING Lab, EPFL

Wouter Lueks
SPRING Lab, EPFL

Julien Pierre Martin
Independent

Soline Ledésert
ICIJ

Anne L'Hôte
ICIJ

Bruno Thomas
ICIJ

Laurent Girod
SPRING Lab, EPFL

Carmela Troncoso
SPRING Lab, EPFL

Abstract

Investigative journalists collect large numbers of digital documents during their investigations. These documents can greatly benefit other journalists' work. However, many of these documents contain sensitive information. Hence, possessing such documents can endanger reporters, their stories, and their sources. Consequently, many documents are used only for single, local, investigations. We present DATASHARENETWORK, a decentralized and privacy-preserving search system that enables journalists worldwide to find documents via a dedicated network of peers. DATASHARENETWORK combines well-known anonymous authentication mechanisms and anonymous communication primitives, a novel asynchronous messaging system, and a novel multi-set private set intersection protocol (MS-PSI) into a *decentralized peer-to-peer private document search engine*. We prove that DATASHARENETWORK is secure; and show using a prototype implementation that it scales to thousands of users and millions of documents.

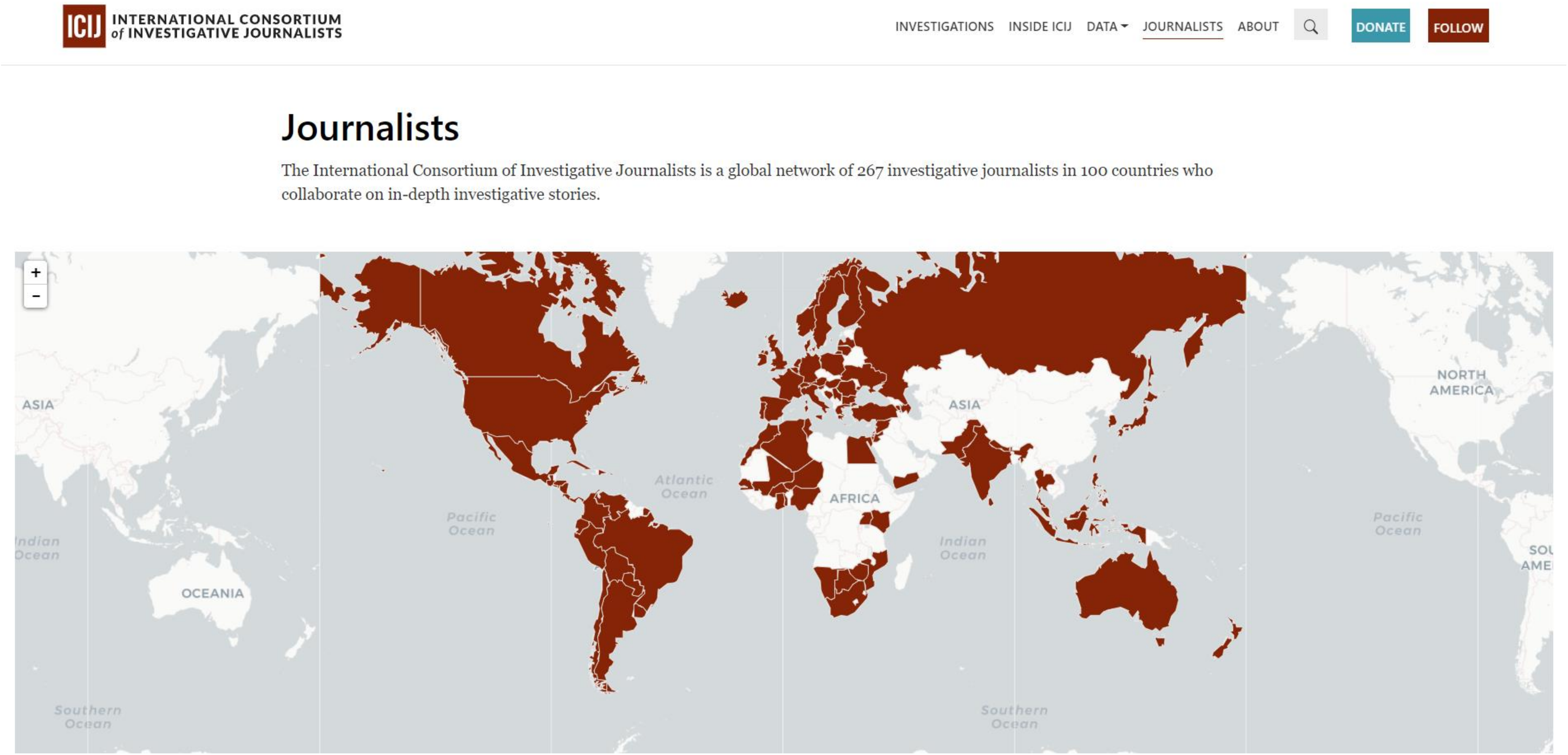
1 Introduction

Investigative journalists research topics such as corruption, crime, and corporate misbehavior. Two well-known examples of investigative projects are the Panama Papers that resulted in several politicians' resignations and sovereign states recovering hundreds of millions of dollars hidden in offshore accounts [27], and the Boston Globe investigation on child abuse that resulted in a global crisis for the Catholic Church [22]. Investigative journalists' investigations are essential for a healthy democracy [10]. They provide the public with information kept secret by governments and corporations. Thus, effectively holding these institutions accountable to society at large.

In order to obtain significant, fact-checked, and impactful results, journalists require large amounts of documents. In a globalized world, local issues are increasingly connected to global phenomena. Hence, journalists' collections can be relevant for other colleagues working on related investigations. However, documents often contain sensitive and/or confidential information and possessing them puts journalists and their sources increasingly at risk of identification, prosecution, and persecution [33,34]. As a result journalists go to great lengths to protect both their documents and their interactions with other journalists [35]. With these risks in mind, the International Consortium of Investigative Journalists (ICIJ) approached us with this question: *Can a global community of journalists search each other's documents while minimizing the risk for them and their sources?*

Building a practical system that addresses this question entails solving five key challenges:
1) *Avoid centralizing information*. A party with access to all the documents and journalists' interaction would become a very tempting target for attacks by hackers or national agencies, and for legal cases and subpoenas by governments.
2) *Avoid reliance on powerful infrastructure*. Although ICIJ has journalists worldwide, it does not have highly available servers in different jurisdictions.
3) *Deal with asynchrony and heterogeneity*. Journalists are spread around the world. There is no guarantee that they are online at the same time, or that they have the same resources.
4) *Practical on commodity hardware*. Journalists must be able to search documents and communicate with other journalists without this affecting their day-to-day work. The system must be efficient both computationally and in communication costs.
5) *Enable data sovereignty*. Journalists are willing to share but not unconditionally. They should be able to make informed decisions on revealing documents, on a case-by-case basis.

The first four requirements preclude the use of existing advanced privacy-preserving search technologies, whereas the fifth requirement precludes the use of automatic and rule-based document retrieval. More concretely, the first requirement prevents the use of central databases and private information retrieval (PIR) [7,23,30] between journalists, as standard PIR requires a central list of all searchable (potentially sensitive) keywords. The second requirement rules out multi-party computation (MPC) between distributed servers [25,40,41].

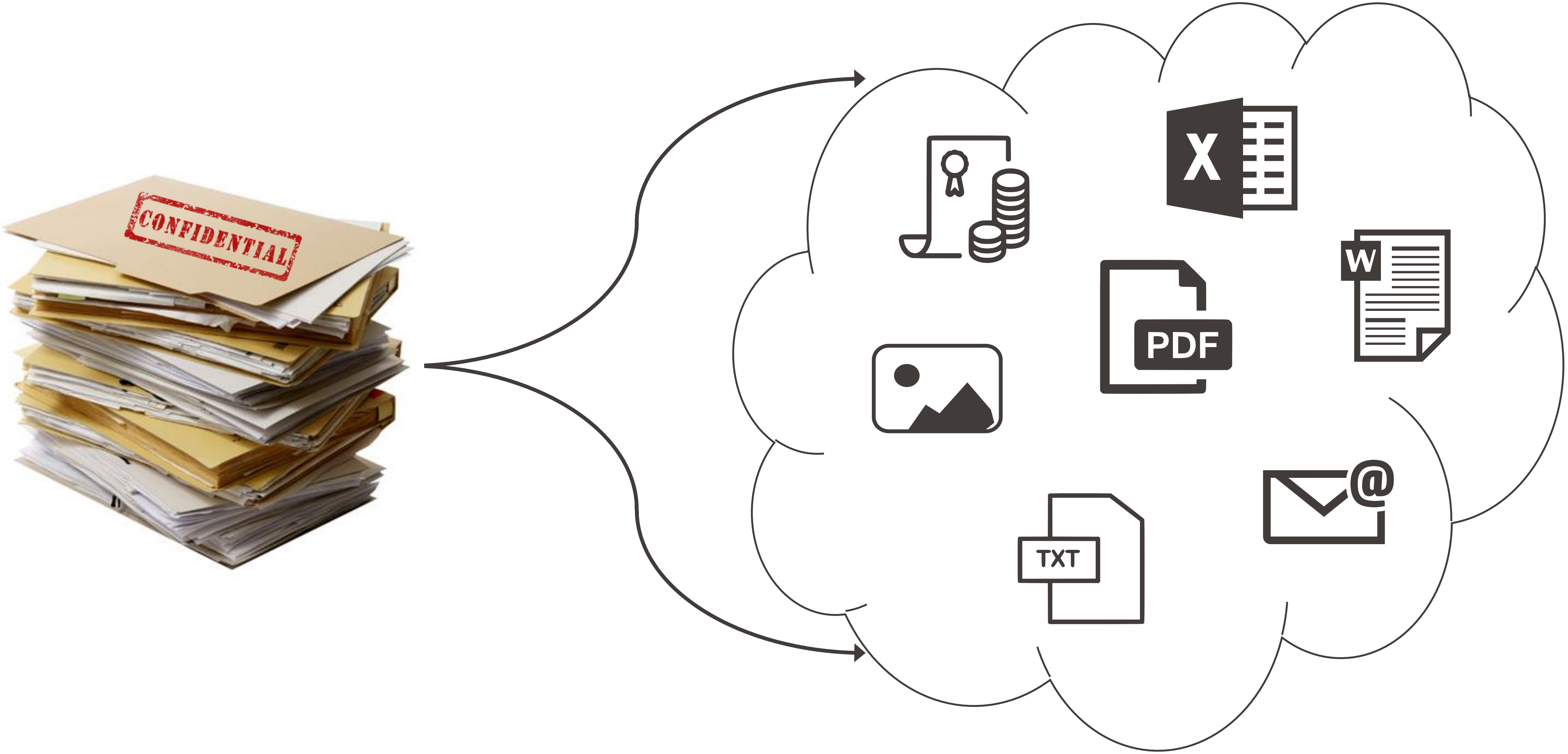


Journalists

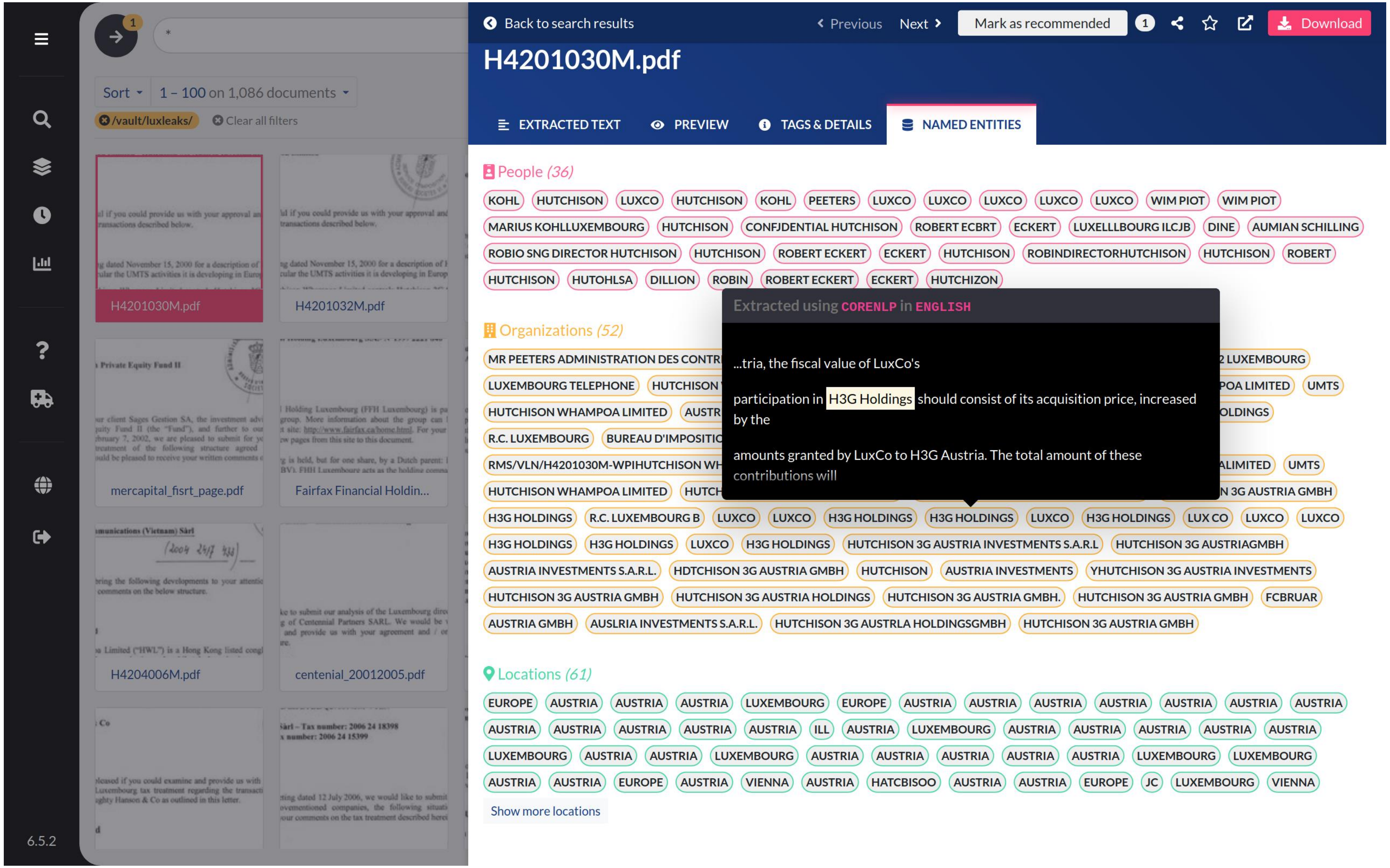
The International Consortium of Investigative Journalists is a global network of 267 investigative journalists in 100 countries who collaborate on in-depth investigative stories.

- https://www.icij.org/

Digitalization



Local document search



H4201030M.pdf

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)

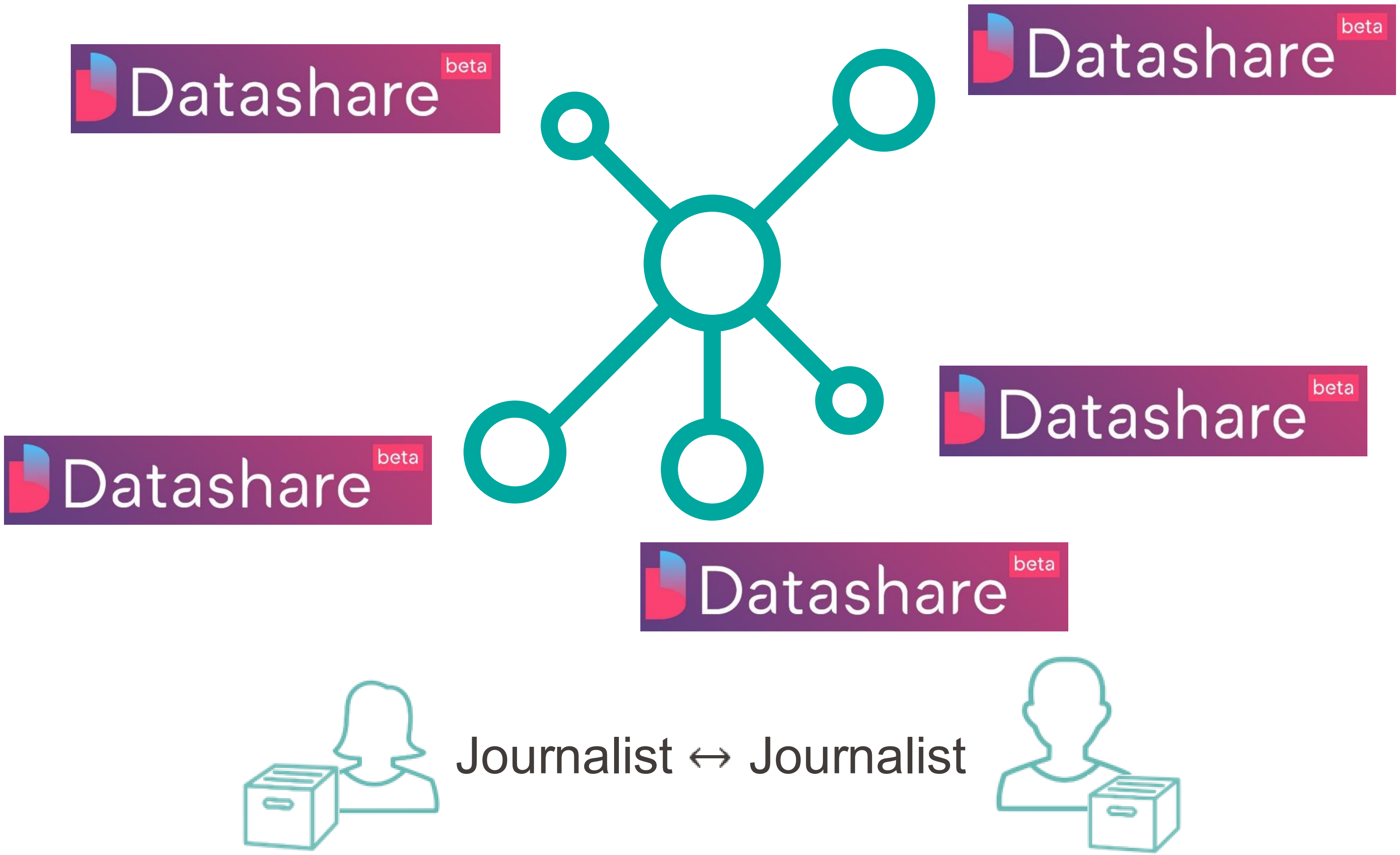
MR PEETERS ADMINISTRATION DES CONTR
2 LUXEMBOURG
LUXEMBOURG TELEPHONE
HUTCHISON
POA LIMITED
UMTS
HUTCHISON WHAMPOA LIMITED
AUSTR
OLDINGS
R.C. LUXEMBOURG
BUREAU D'IMPOSITIO
RMS/VLN/H4201030M-WPIHUTCHISON WH
ALIMITED
UMTS
HUTCHISON WHAMPOA LIMITED
HUTCH
N 3G AUSTRIA GMBH
H3G HOLDINGS
R.C. LUXEMBOURG B
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

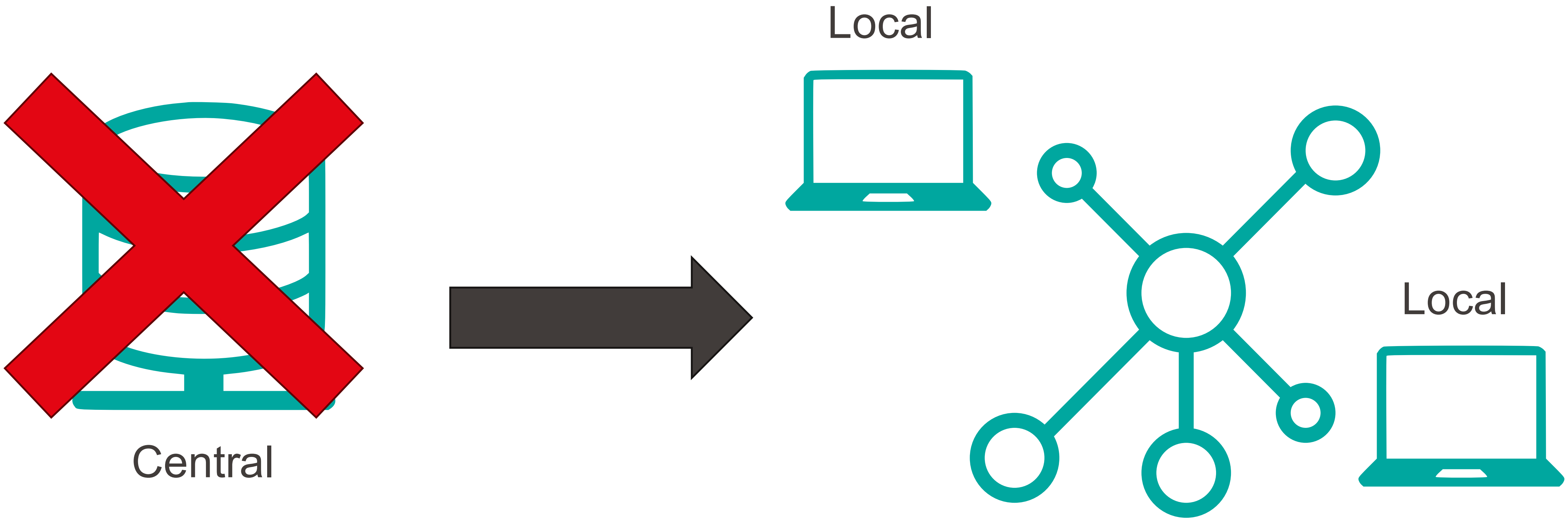
Show more locations

The goal: a decentralized search engine



Journalist ↔ Journalist

First:
A (not so)clear goal



Central

Local

Local

Journalist ↔ Journalist

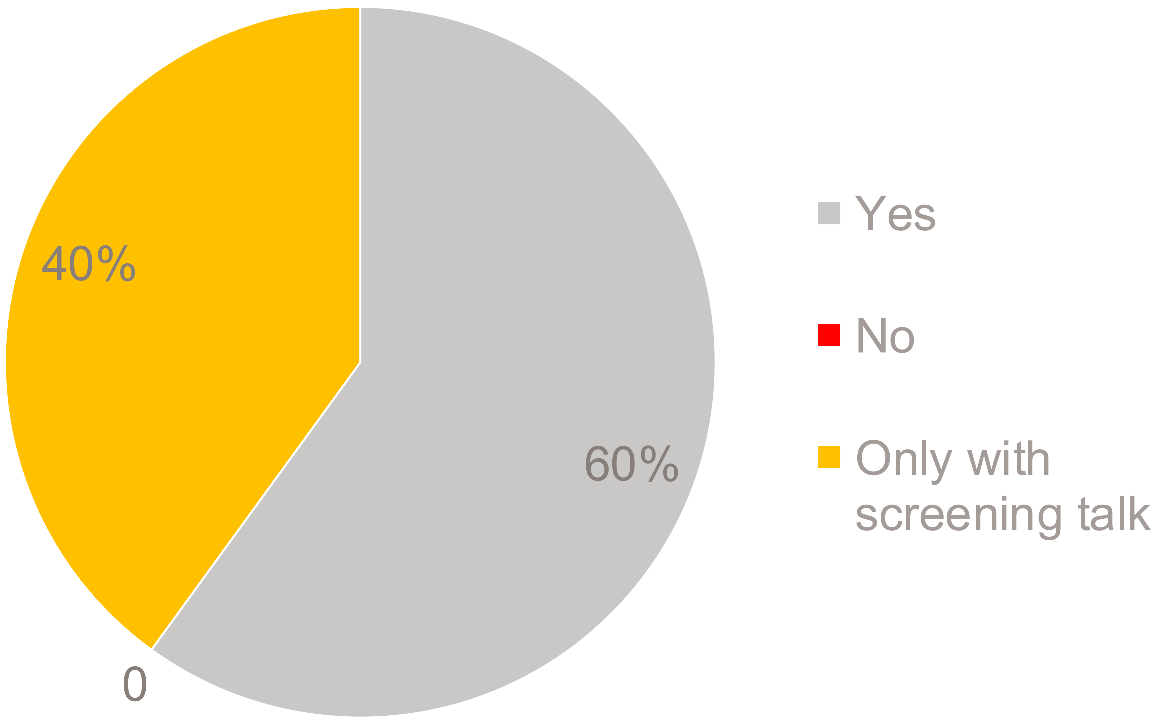
Clarifying the goal

- **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?



Datashare Network

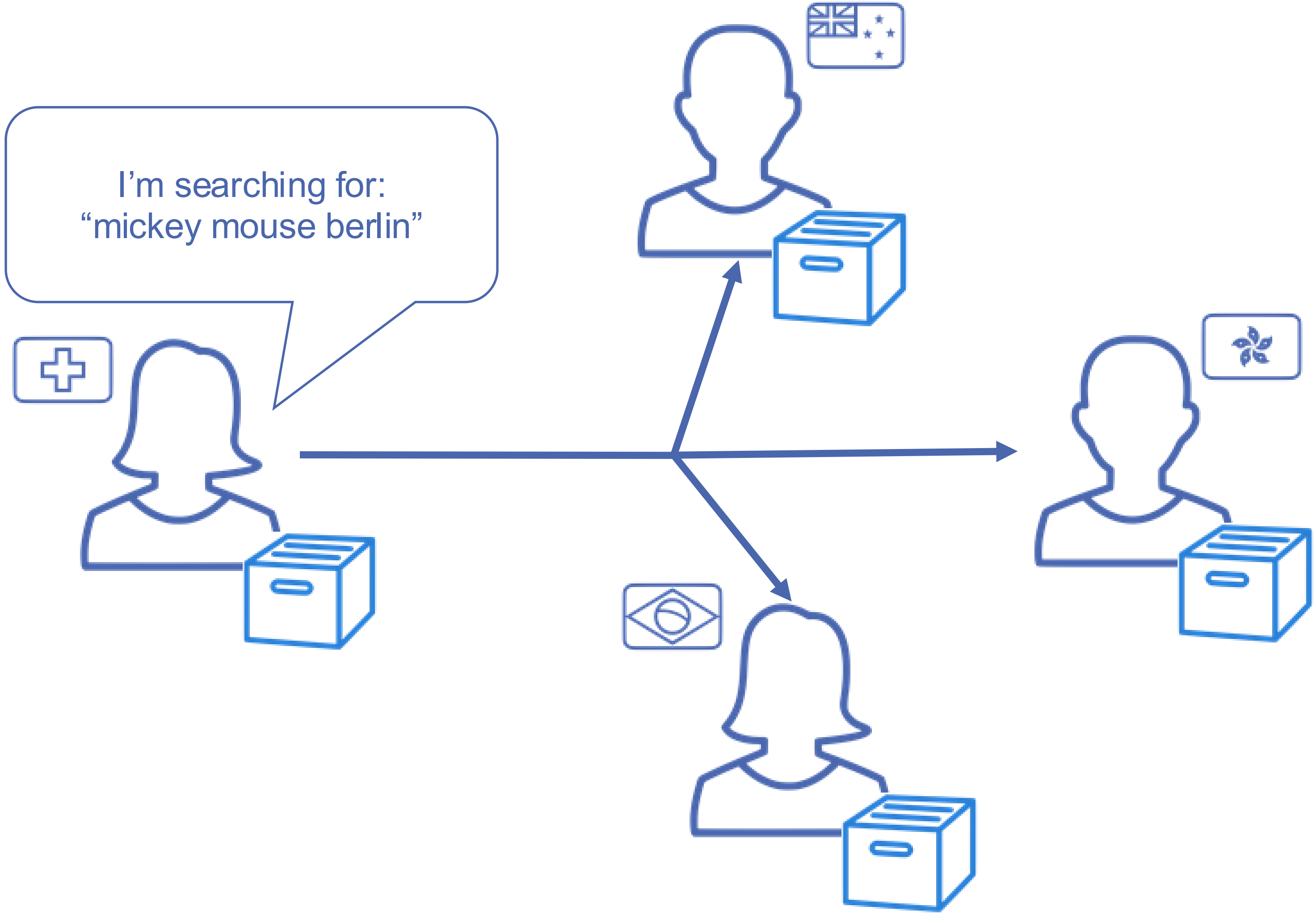
Enable journalists to search on others' collections for keywords of interest.

Protect journalists & sources.

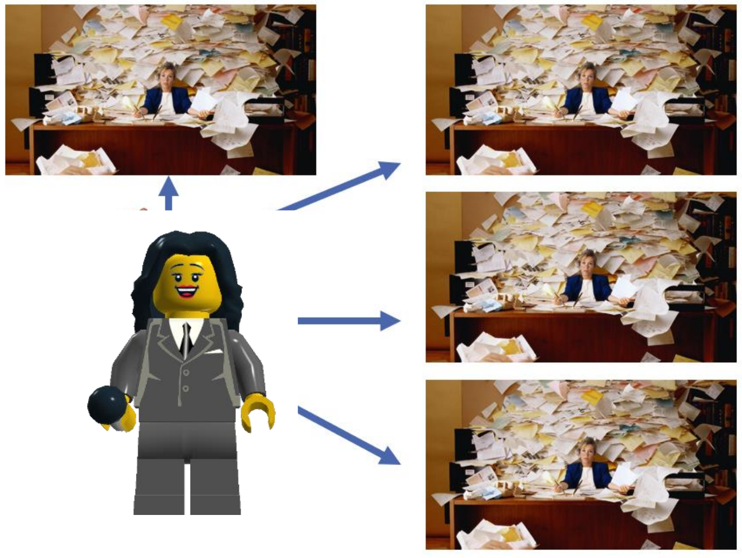
Only ICIJ and associates can use the system.

No one (journalists, ICIJ, others) can learn:

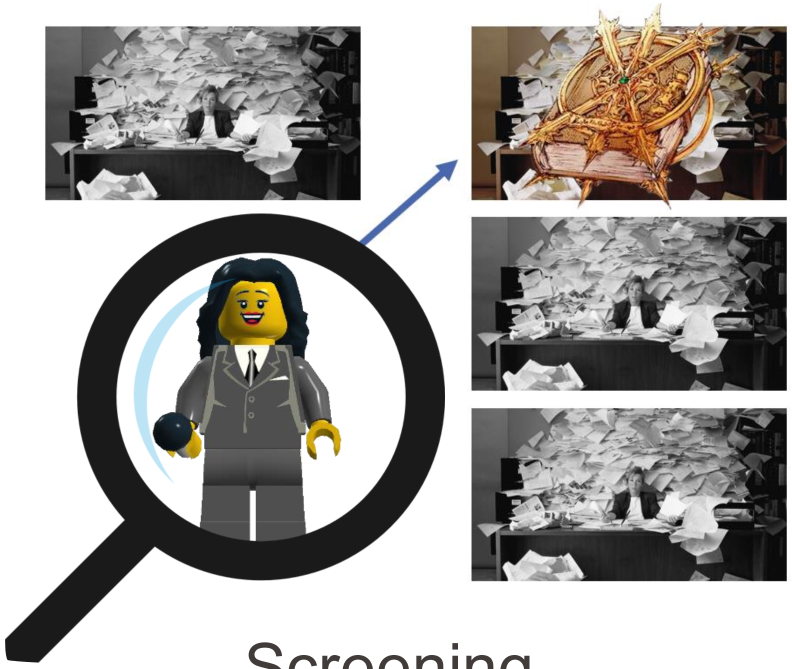
- **who** queries
- **what** is queried
- **whole** document collections



Required functionality

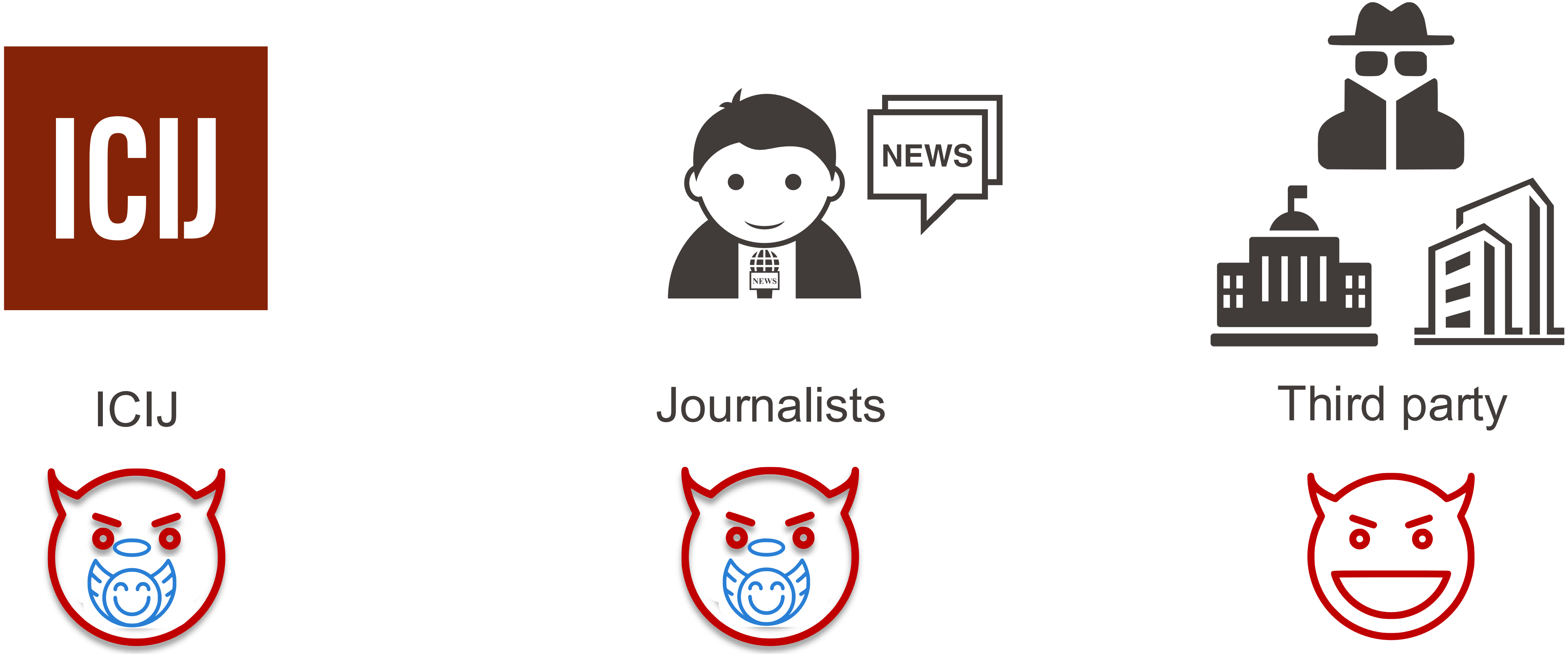


Search



Screening

Security and Privacy Requirements



ICIJ

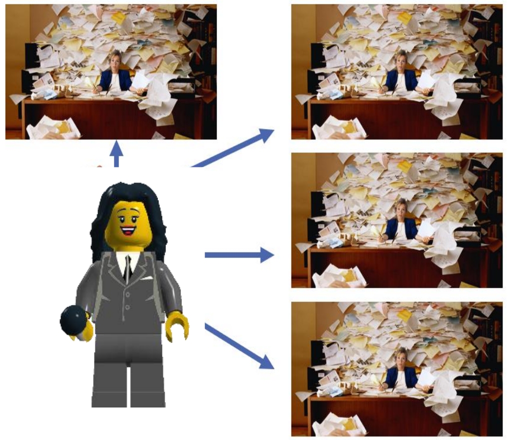
Journalists

Third party

Real-world constraints

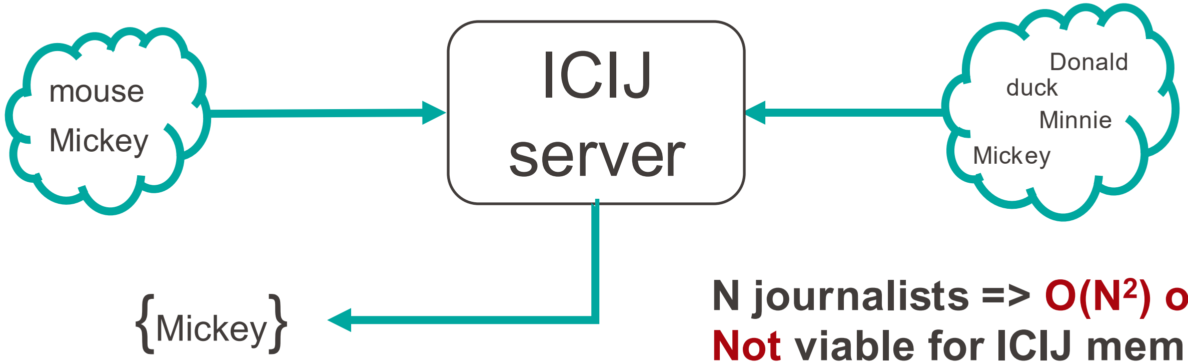
- Asynchrony
- Scarce resources
  - **Computation**
  - **Bandwidth**
- But… no real time or infrastructural requirements

Required functionality



Search

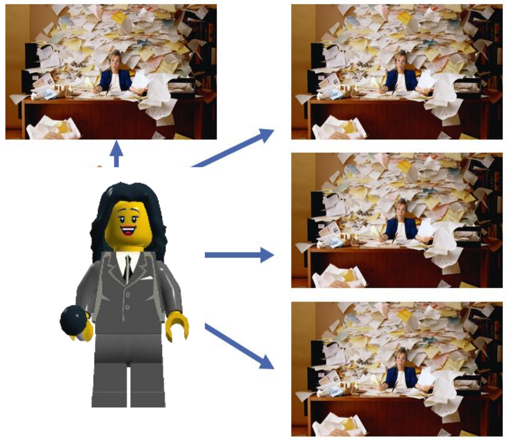
Existing: Private Set Intersection



N journalists => O(N^2) operations

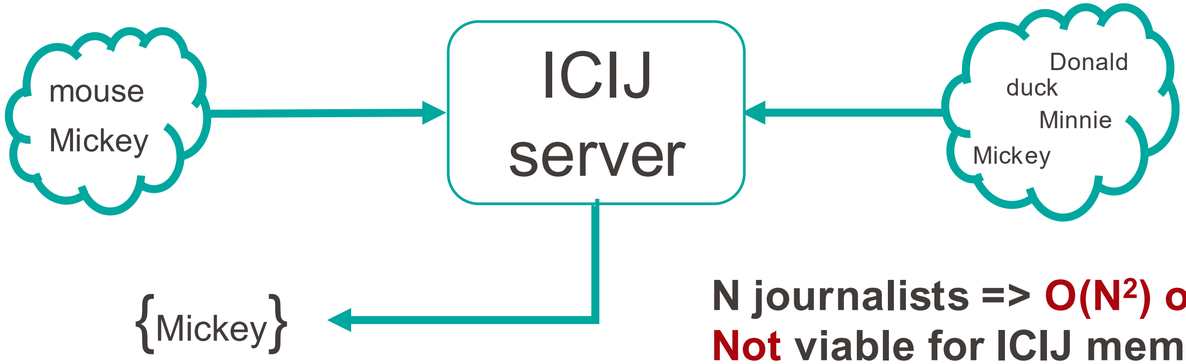
Not viable for ICIJ members

Required functionality



Search

Existing: Private Set Intersection



mouse
Mickey

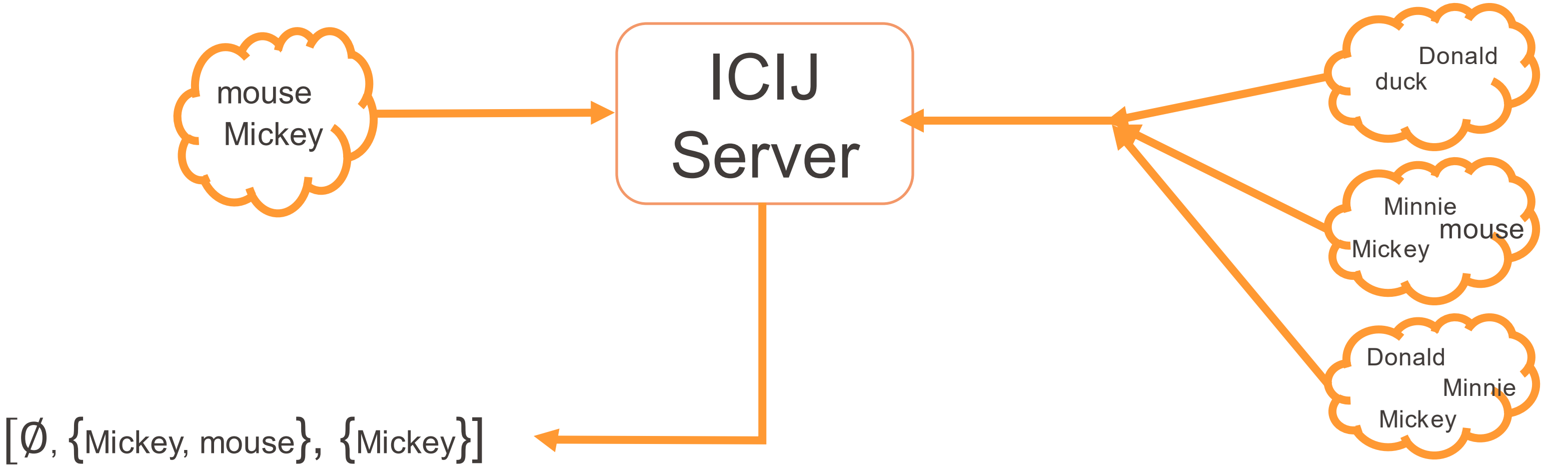
ICIJ server

Donald
duck
Minnie
Mickey

{Mickey}

N journalists => O(N^2) operations
Not viable for ICIJ members

Our contribution: Multi-set Private Set Intersection



mouse
Mickey

ICIJ Server

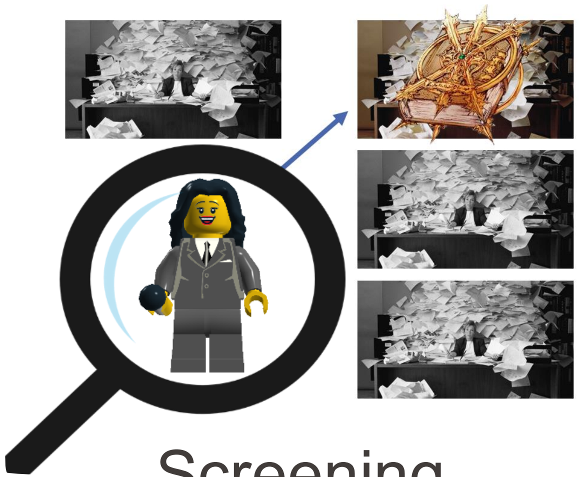
Donald
duck

Minnie
mouse
Mickey

Donald
Minnie
Mickey

[∅, {Mickey, mouse}, {Mickey}]

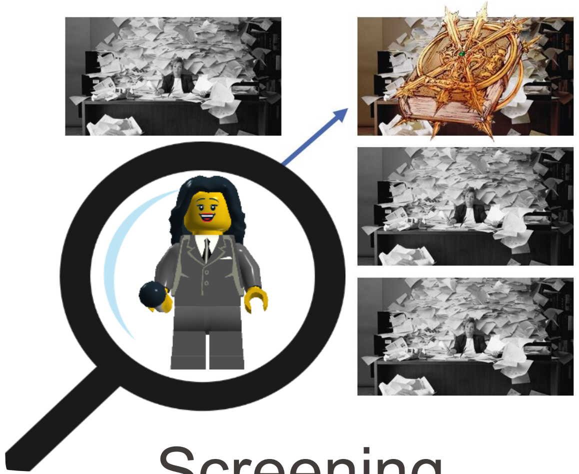
Required functionality



Screening

Existing: No private ephemeral communication system

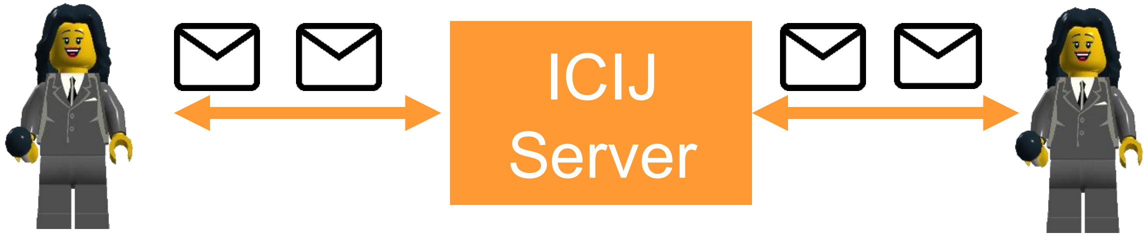
Required functionality



Screening

Existing: No private ephemeral communication system

Our contribution



ICIJ Server

Cryptography-based ephemeral mailboxes
+
Dummy traffic or PIR-based (different trade-off)
+
Anonymous communications

Engineering: Putting it all together

Authentication

Only ICIJ and associates can use the system → Attribute-based credentials

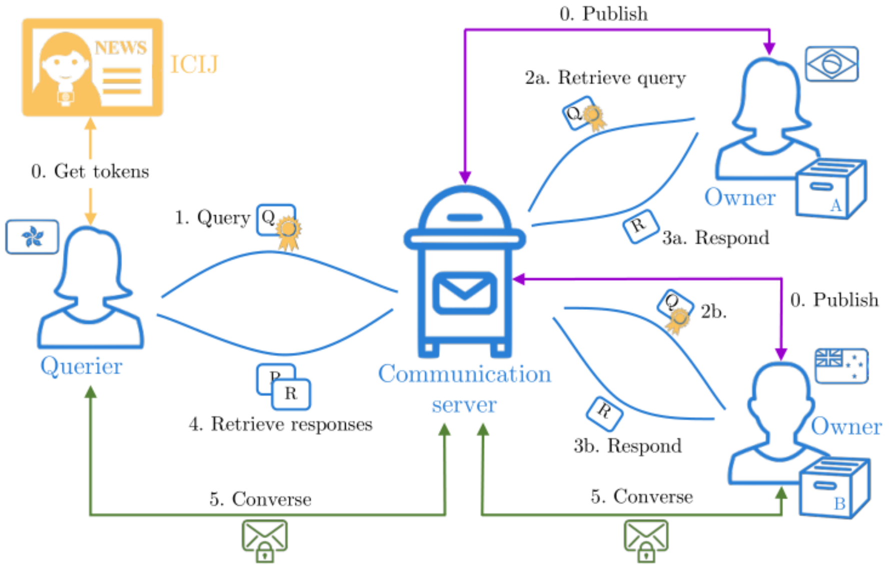
Search

Query content is not revealed → Multi-set private set intersection

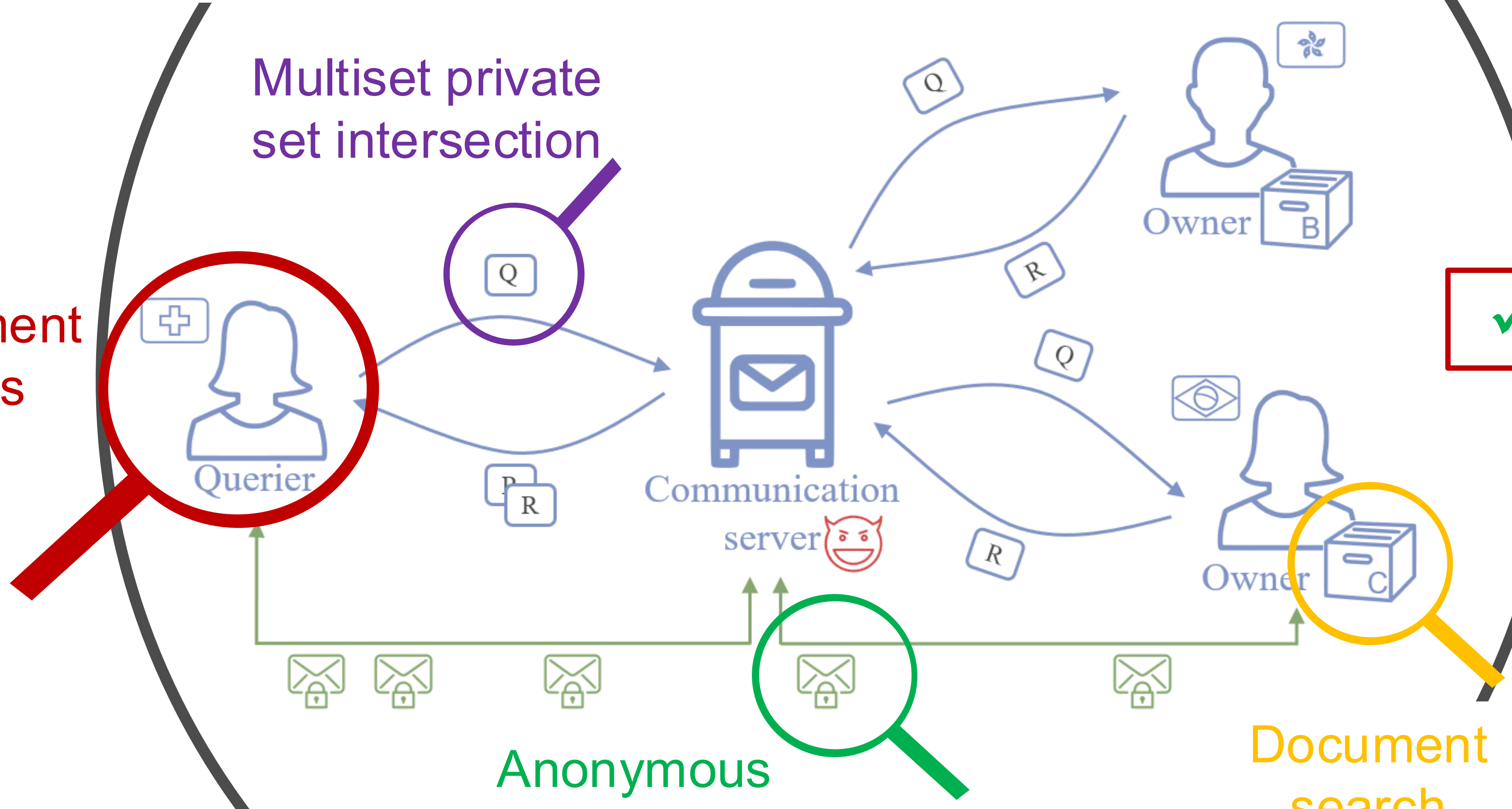
Who searches is anonymous → Anonymous communications

Screening

Anonymous screening talks → Ephemeral mailboxes



Datashare Network
End-to-end privacy engineering



✓ Formal proofs

ICIJ

Take-aways

- Privacy engineering is about limiting harms, via limiting purposes
- Once purpose is identified strategies exist to minimize trust in system entities
- Strategies are implemented by the PETs you have seen throughout the course!
- Combining is hard: quantification is difficult

What if the technologies are not ready?

Or have drawbacks, or cannot fulfill all regulation requirements, or cannot be extended, ...

Take the ideal privacy-by-design system,
and use as a reference for feasible system evaluation.
The feasible system:

- constrains to the same purpose?
- collects more data?
- gives more data to more entities?
- increases the amount of trust?
- who is affected by compromise?