University of Kent | Institute of Cyber Security for Society (iCSS)

Privacy through the Lens of Data Flows

Shujun LI (李树钧)

Director, Institute of Cyber Security for Society (iCSS)

Professor of Cyber Security, School of Computing
University of Kent, UK

http://www.hooklee.com/

@hooklee

Privacy through lens of data flows

University of Kent | Institute of Cyber Security for Society (iCSS)

Privacy in the cyber-physical world

Data on the Internet/Web

University of Kent | Institute of Cyber Security for Society (iCSS)

- Information about your machine / operating system
  - IP addresses, time zone, language, screen size, fonts installed, …
- Sensor data made accessible via web browsers
  - GPS coordinates, camera data, microphone data, …
- Information stored within or by web browsers
  - Web browser type and version, web cookies, web browsing histories and caches, search keywords on search engines, auto completion data, bookmarks, account details (e.g., Google account for Google Chrome), web browser extensions installed, …
- Data stored in web browser extensions
  - May or may not be in the web browser (can be stored online)
- **User-generated content (UGC)**
  - Search queries, online profiles, online posts, …
- …

What online services?

University of Kent | Institute of Cyber Security for Society (iCSS)

- Online social networks (OSN) or social media (SM)
- Microblogging systems
- Online chat rooms
- Instant messaging systems
- Web forums
- Image and video sharing portals
- Search engines
- **Any** websites or web-based applications and (mobile) apps supporting UGCs
  - App marketplaces, online shopping, online booking, news outlets, online dating, e-petition, online maps, …
- …

What UGC?

University of Kent | Institute of Cyber Security for Society (iCSS)

- Profile data
  - Mostly personal data: name(s), facial image (profile image), addresses, geo-location, gender, age, profession, personal interests, financial status, information of family members, etc.
- Online posts, comments, reviews, ratings, …
  - They may contain (a lot of!) personal data.
  - And they are normally linked to profile data.
- …
- Personal and sensitive data of **yourself** and **other people**
  - Neither you nor other people may have noticed you did it.
  - If you disclosed other people's personal data and they saw it, they may not feel comfortable to (publicly) object to your disclosure.
  - You can't control what others will share about you!

Your data may not be just yours!

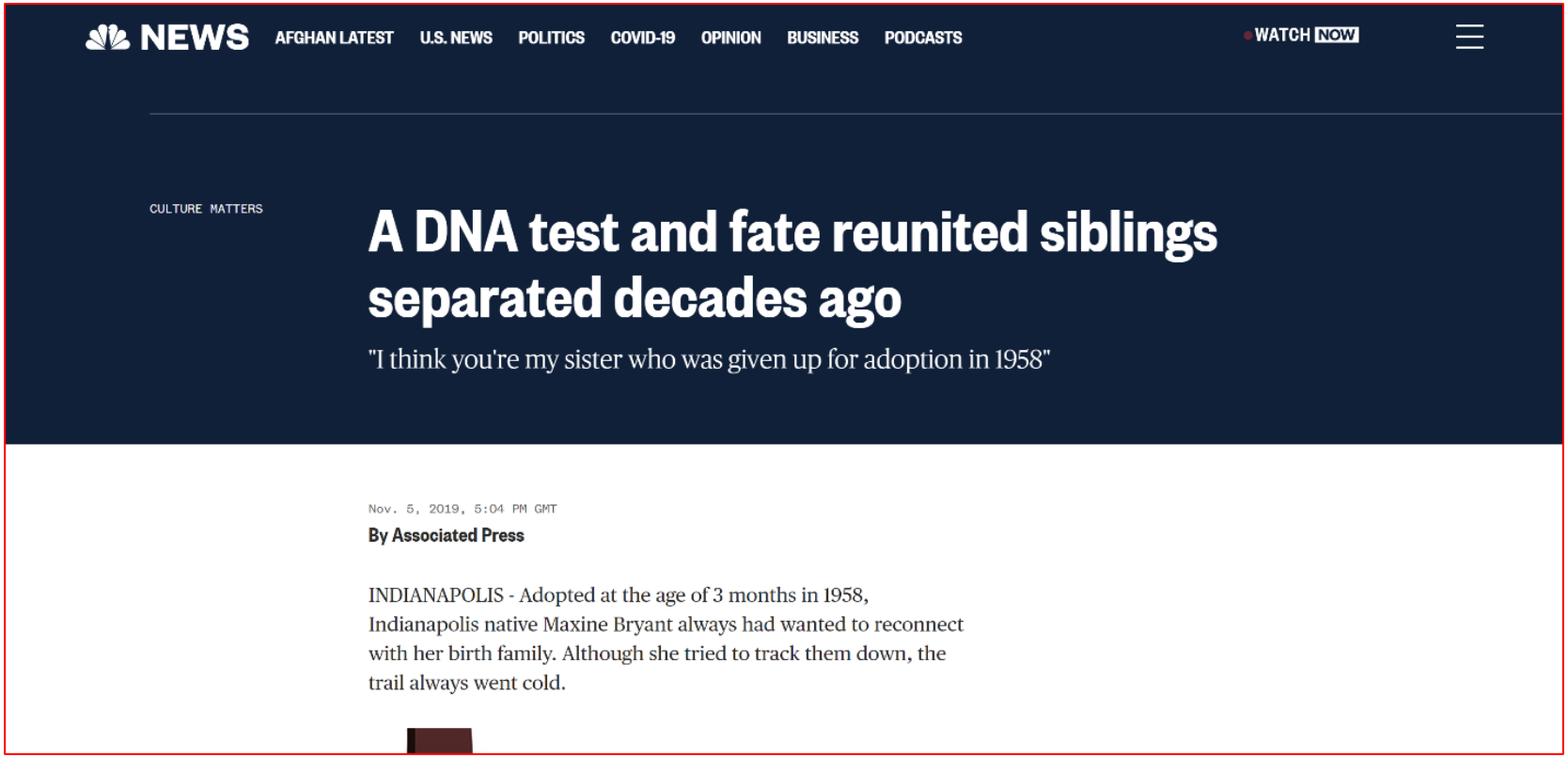
University of Kent | Institute of Cyber Security for Society (iCSS)

- Example: Your (sur)name(s) and your DNA data are your whole **extended** family's, …

Your data may not be just yours!

University of Kent | Institute of Cyber Security for Society (iCSS)

- Example: Your DNA data is your whole **extended** family's, including "unknown" relatives.

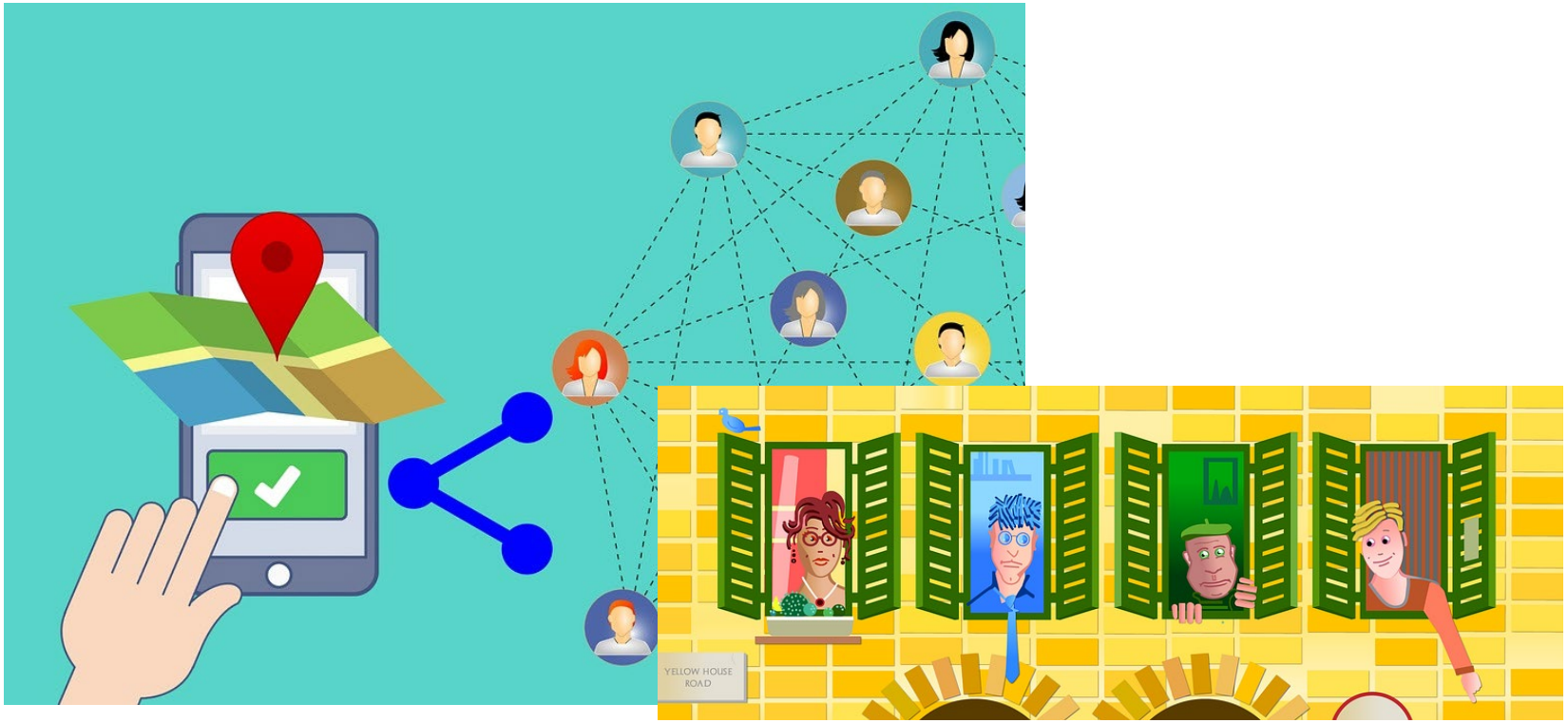


Source: https://www.nbcnews.com/news/nbcblk/dna-test-fate-reunited-siblings-separated-decades-ago-n1076006

Your data may not be just yours!

University of Kent | Institute of Cyber Security for Society (iCSS)

- Example: Your geo-location, life patterns, religion, cultural background also belong to your neighbours, friends and those living nearby!



Where are the data?

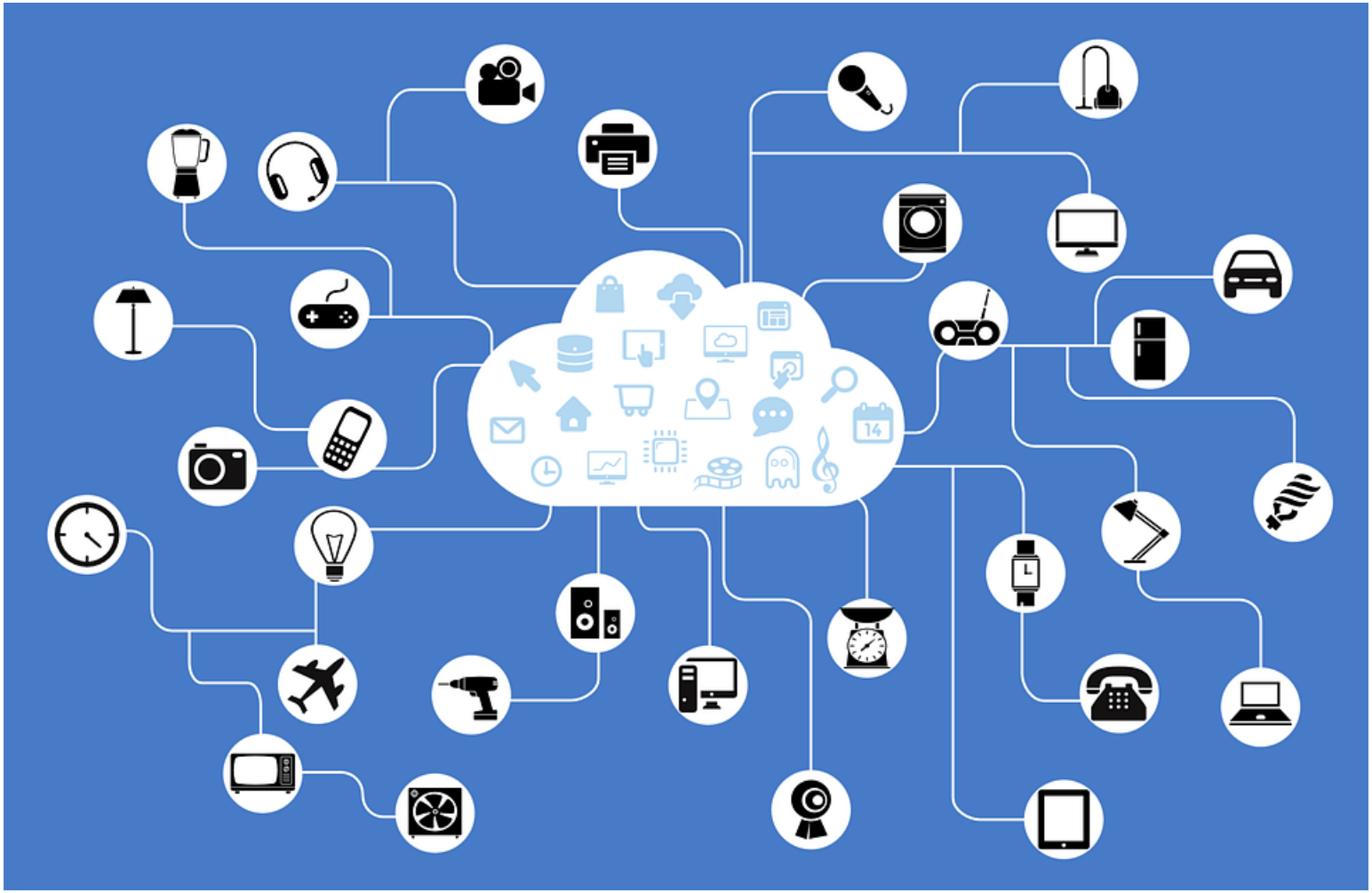
University of Kent | Institute of Cyber Security for Society (iCSS)

- On your devices or on the Internet/Web?
- Where exactly?

Where are the data?

University of Kent | Institute of Cyber Security for Society (iCSS)

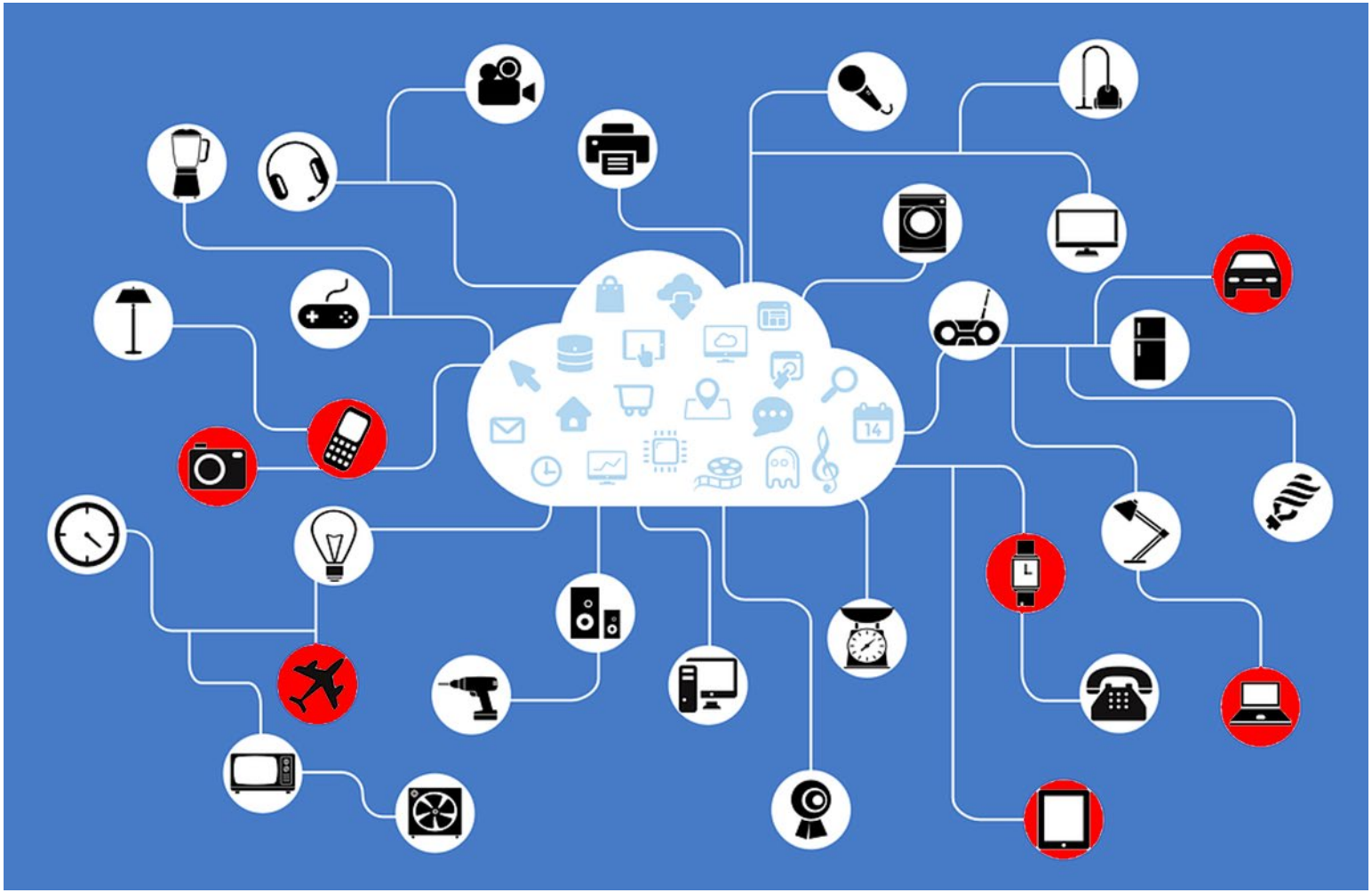
- IoT everywhere? Where is the cloud?



Where are the data?

University of Kent | Institute of Cyber Security for Society (iCSS)

- At your home/office or while travelling?



Example: CCTV and smart cities

University of Kent | Institute of Cyber Security for Society (iCSS)

- CCTV and cameras everywhere?

Example: Privacy in public spaces

University of Kent | Institute of Cyber Security for Society (iCSS)

Global Voices advox

Newsletter ▾ About Us ▾ Search

A project of **Global Voices**, we are a global anti-censorship network of bloggers and activists dedicated to protecting freedom of expression online.

Activism Privacy Netizen Report About Our Community Archives

The Russian Art of Meta-Stalking

Posted 7 April 2016 21:38 GMT

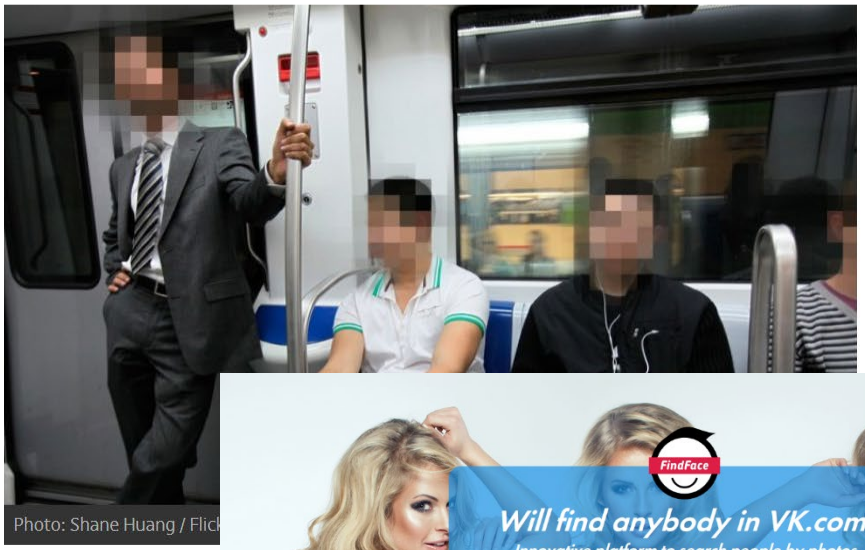
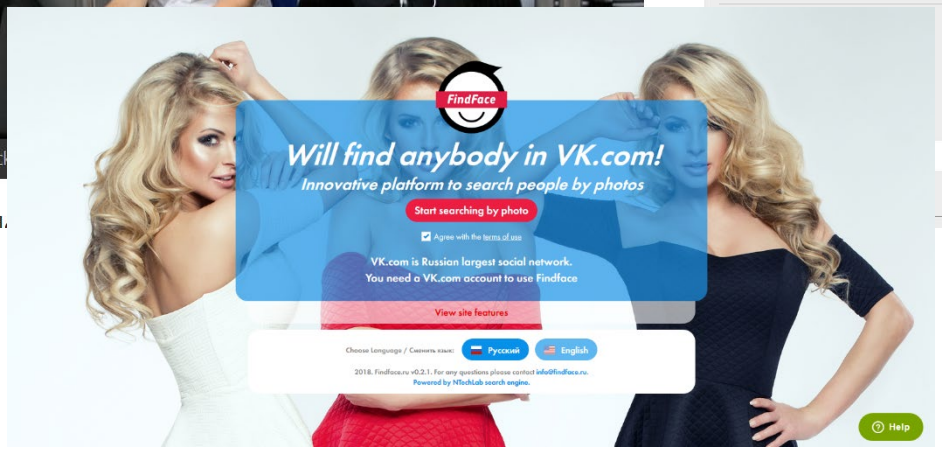


Photo: Shane Huang / Flick



Written by
Kevin Rothrock

Read this post in Ελληνικά, Español, English

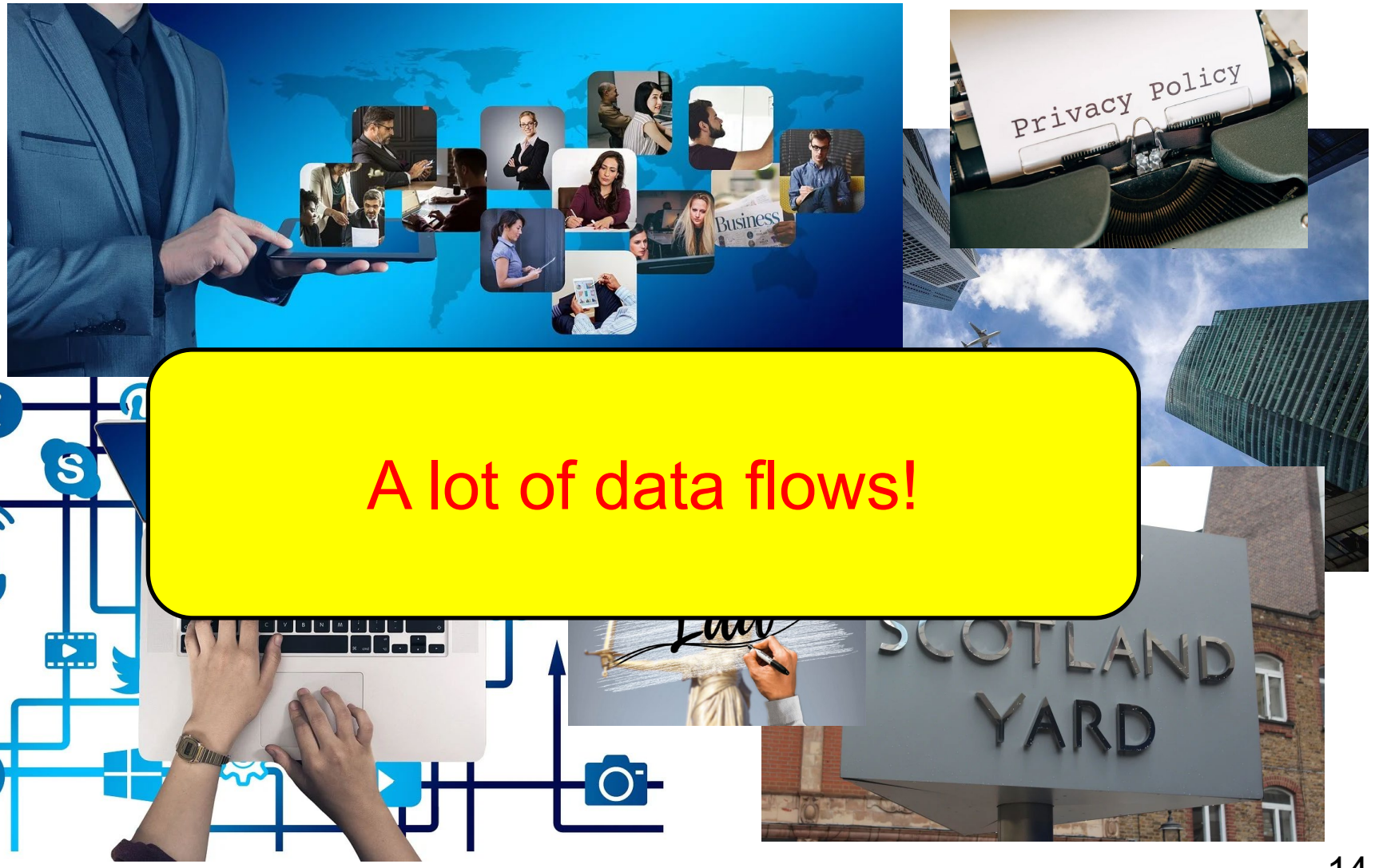
Harassment, Media, Tech & Tools

Russia

Eastern & Central Europe

Who are data consumers and why?

University of Kent | Institute of Cyber Security for Society (iCSS)



A lot of data flows!

Privacy through lens of data flows

University of Kent | Institute of Cyber Security for Society (iCSS)

The  PRIVELT Project

Basic information

PRiVELT

University of Kent | Institute of Cyber Security for Society (iCSS)

- Title: PRIvacy-aware personal data management and Value Enhancement for Leisure Travellers (**PriVELT**)
- Funder: 

- Call: Trust, Identity, Privacy and Security in the Digital Economy 2.0 (2018)
- Budget: £~1.4m
- Duration: 10/2018 – 06/2023 (57 months)
- Website: https://privelt.ac.uk/

University of Kent

UNIVERSITY OF SURREY

NTU Nottingham Trent University

WARWICK
THE UNIVERSITY OF WARWICK

Durham University

(Part of) The (former) project team

University of Kent | Institute of Cyber Security for Society (iCSS)

Yang Lu
Research Associate

Athina Ioannou
Research Associate

Shujun Li (Consortium PI)
Professor of Cyber Security

Iis Tussyadiah (Partner PI)
Professor in Hospitality & Digital Experience

1. Computer Science
2. Tourism
3. Business
4. Psychology
5. Law

Alex Freitas (Co-I)
Professor of Computational Intelligence

Graham Miller (Co-I)
Professor of Sustainability in Business

Lisa Dickson (Co-I)
Reader in Law

Annabelle Gawer (Co-I)
Professor of Digital Economy

Mario Weick (Co-I)
Associate Professor in Quantitative Social Psychology

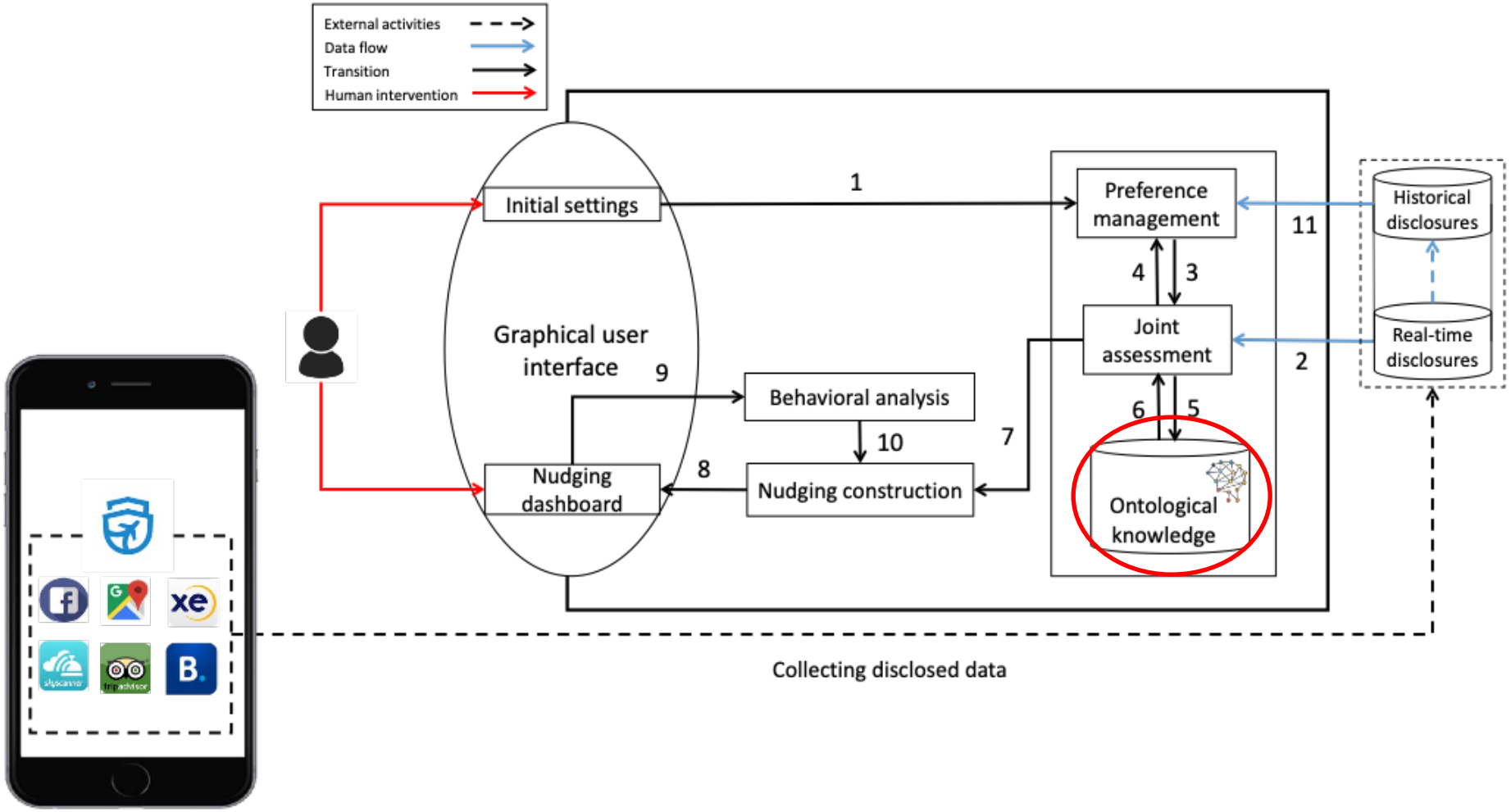
Research Associate
Diwali Diwali

Jay Bal (Partner PI)
Professor, WMG

Xiao Ma (Co-I)
Associate Professor, WMG

The vision: user-centric, server-less, from privacy awareness to nudging

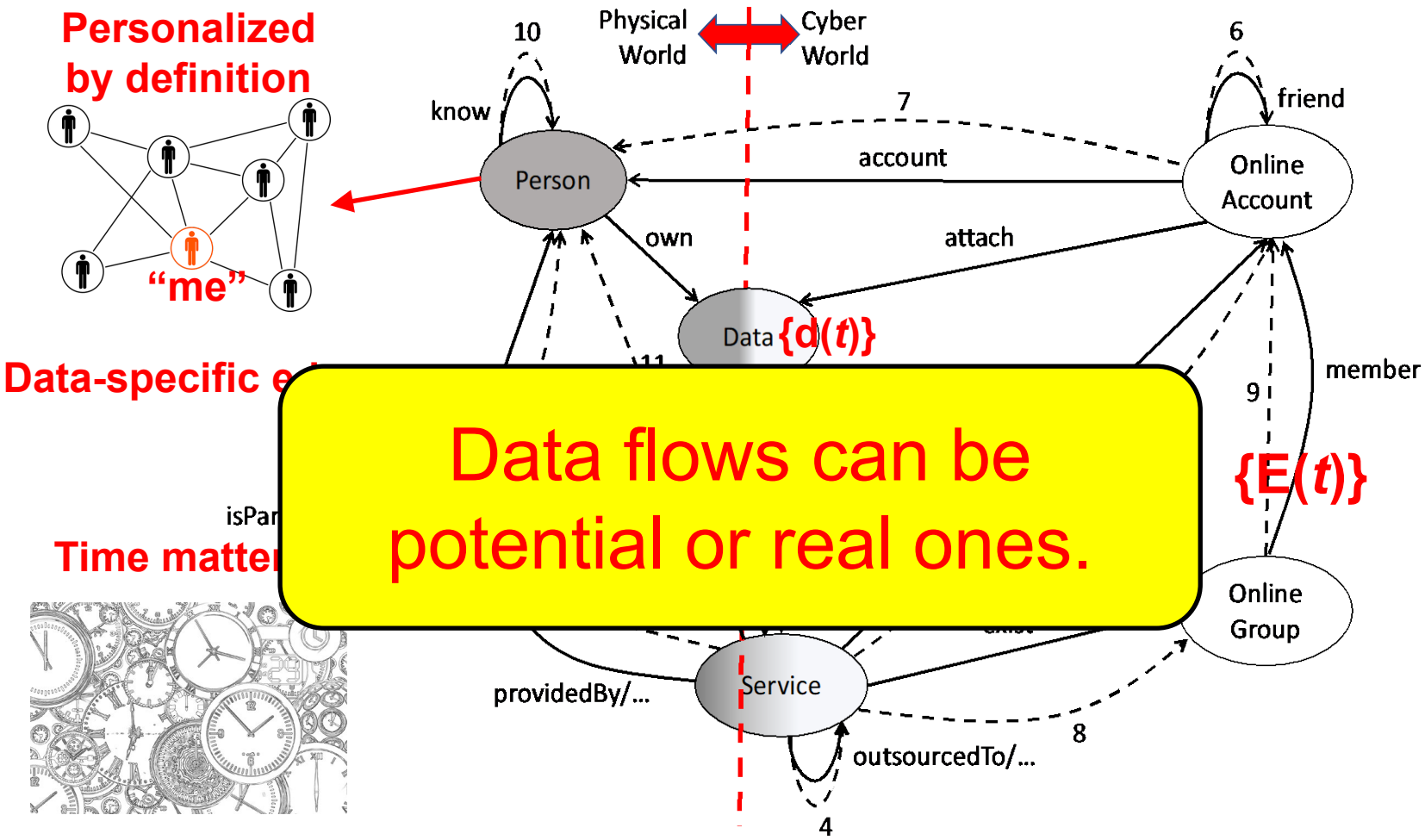
University of Kent | Institute of Cyber Security for Society (iCSS)



Yang Lu, Shujun Li, Athina Ioannou and Iis Tussyadiah (2019) From Data Disclosure to Privacy Nudges: A Privacy-aware and User-centric Personal Data Management Framework. In *Proc. DependSys 2019*, Springer. doi:10.1007/978-981-15-1304-6_21

Data sharing (flow) ontology

University of Kent | Institute of Cyber Security for Society (iCSS)



Yang Lu and Shujun Li (2020) From Data Flows to Privacy Issues: A User-Centric Semantic Model for Representing and Discovering Privacy Issues. In *Proc. HICSS 2020*, University of Hawai'i at Mānoa. doi: 10.24251/HICSS.2021.651

Is a data flow graph complex?

University of Kent | Institute of Cyber Security for Society (iCSS)

- Number of nodes: **large**
  - "Me": the "center" / owner of the graph
  - All data item and data packages about "me"
  - All people your data can flow to (could be **anyone**)
  - All physical and online services you data can flow to
  - All organizations your fata can flow to
  - *Exercise: Check your password manager!*
- Number of edges: **huge**
  - Relationships between different types of nodes
  - Often more than one edge between any two nodes
  - *Exercise: Check the data dashboard of your Google Account or your account with any other online service!*

Useful metrics of a data flow graph

University of Kent | Institute of Cyber Security for Society (iCSS)

- Out-degree (of “me” node), given a time window
  - The amount of data shared
- Average nodal degree (of a data consumer node)
  - The average amount of “my” data disclosed to that data comnsumer
- Node / Link connectivity (of the whole graph)
  - The number of “essential” data consumers / data sharing activities
- Centrality metrics (of data consumer nodes)
  - For identifying major (potentially “hidden”) data consumers
- The longest path(s) originating from “me”
  - For identifying the most “hidden” data consumer(s)
- Network type (of the whole graph)
  - Small-world network, scale-free network or something else?
- …

“Topological” privacy issues

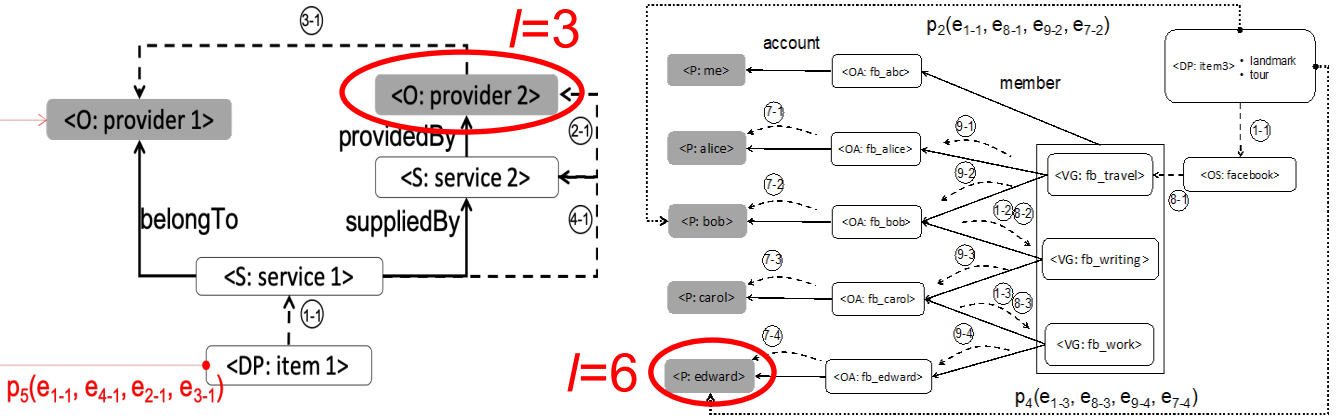
University of Kent | Institute of Cyber Security for Society (iCSS)

- A specific privacy issue with a specific data item or a data package corresponds to a **data flow path**.
- A specific privacy issue with more than one data items and/or data packages corresponds to **a set of data flow paths**.
- A specific **type** of privacy issues of one or more data item / package type(s) is **a set of sets of data flow paths**.
- All of them can be **described** and **potentially detected** via their **common topological features**.

“Topological” privacy issue #1

University of Kent | Institute of Cyber Security for Society (iCSS)

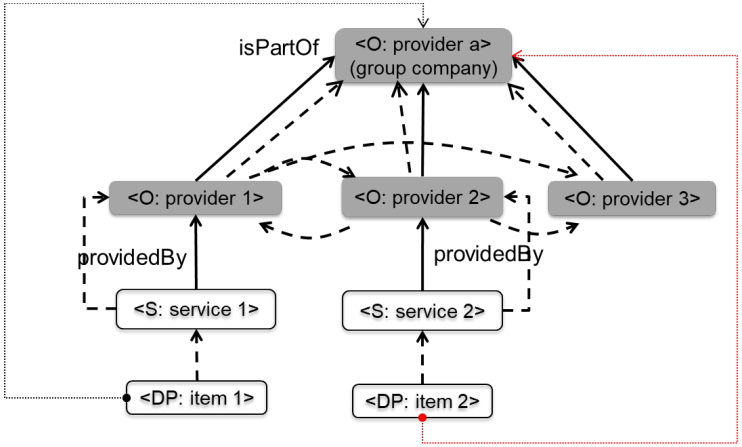
- Data shared with (potentially) **unknown consumers**
  - **Hypothesis**: the longer the data flow path length between “me” and a data consumer node is, the more likely the user is unaware of the data consumer
  - **Risk assessment**: $r=f(l)$, where l is the path length
  - **Visualization**: show a ranked list of all potential unknown data consumers with decreasing values of r
  - **Detection (naïve method)**: $r > r_t \Rightarrow$ issue an alert



“Topological” privacy issue #2

University of Kent | Institute of Cyber Security for Society (iCSS)

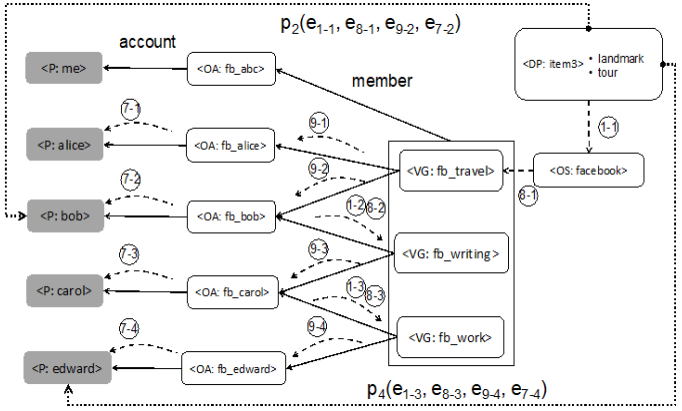
- Indirect (= potentially unknown) **data aggregator**
  - **Hypothesis**: given a tree whose root node is a data consumer, the taller the tree is, the more likely the root node is an unknown (super) data aggregator
  - **Risk assessment**: $r=f(h)$, where h is the tree’s height
  - **Visualization**: show a ranked list of all potential data aggregators with decreasing values of r
  - **Detection (naïve method)**: $r > r_t \Rightarrow$ issue an alert



“Topological” privacy issue #3

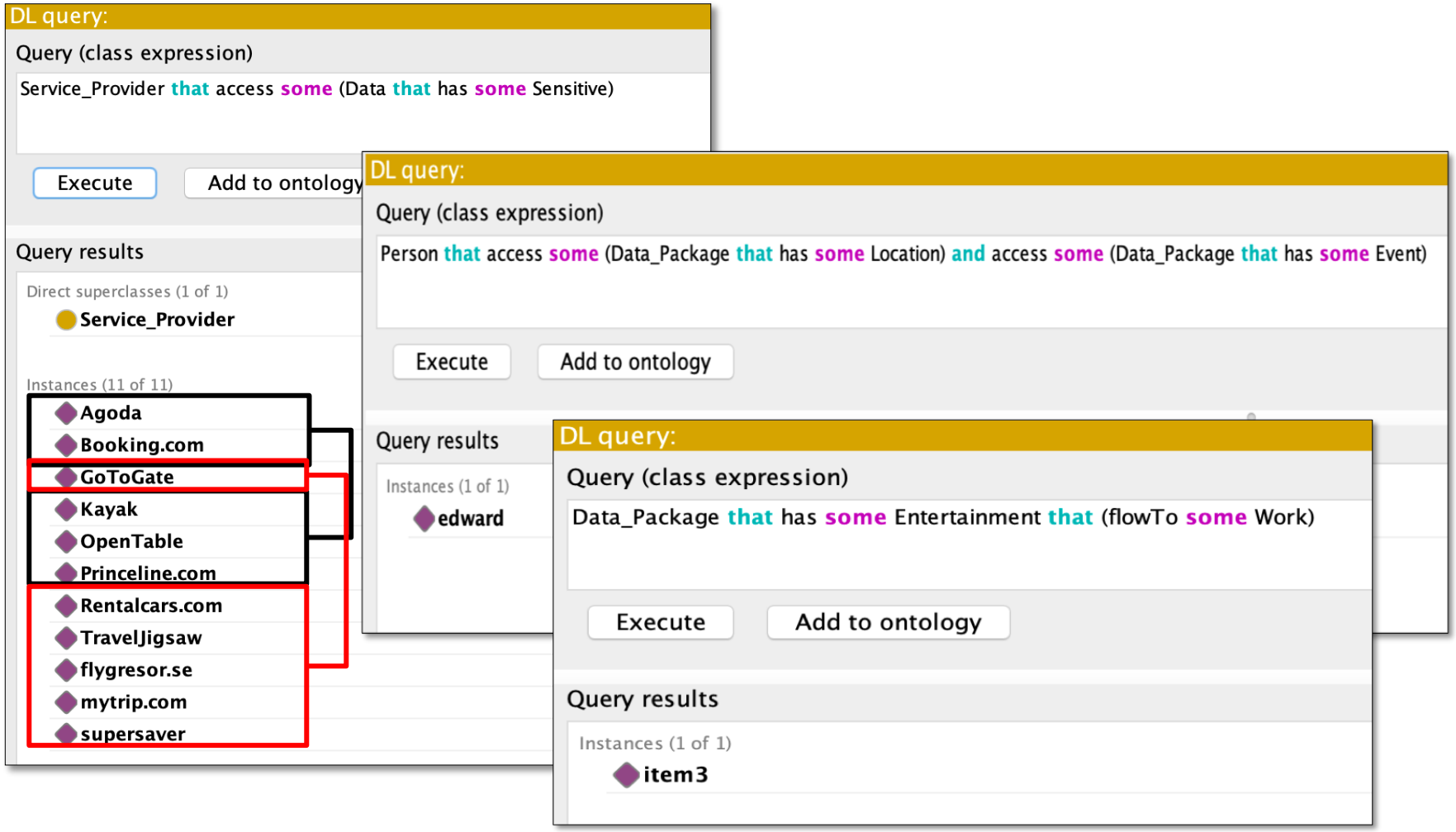
University of Kent | Institute of Cyber Security for Society (iCSS)

- Data shared with **too many consumers**
  - **Hypothesis**: given a tree whose root is a data node, the bigger the tree is, the more likely the data has been over-shared too much
  - **Risk assessment**: $r=f(n)$, where n is the total number of nodes in the tree minus 1 (the root node)
  - **Visualization**: show the whole tree
  - **Detection (naïve method)**: $r > r_t \Rightarrow$ issue an alert



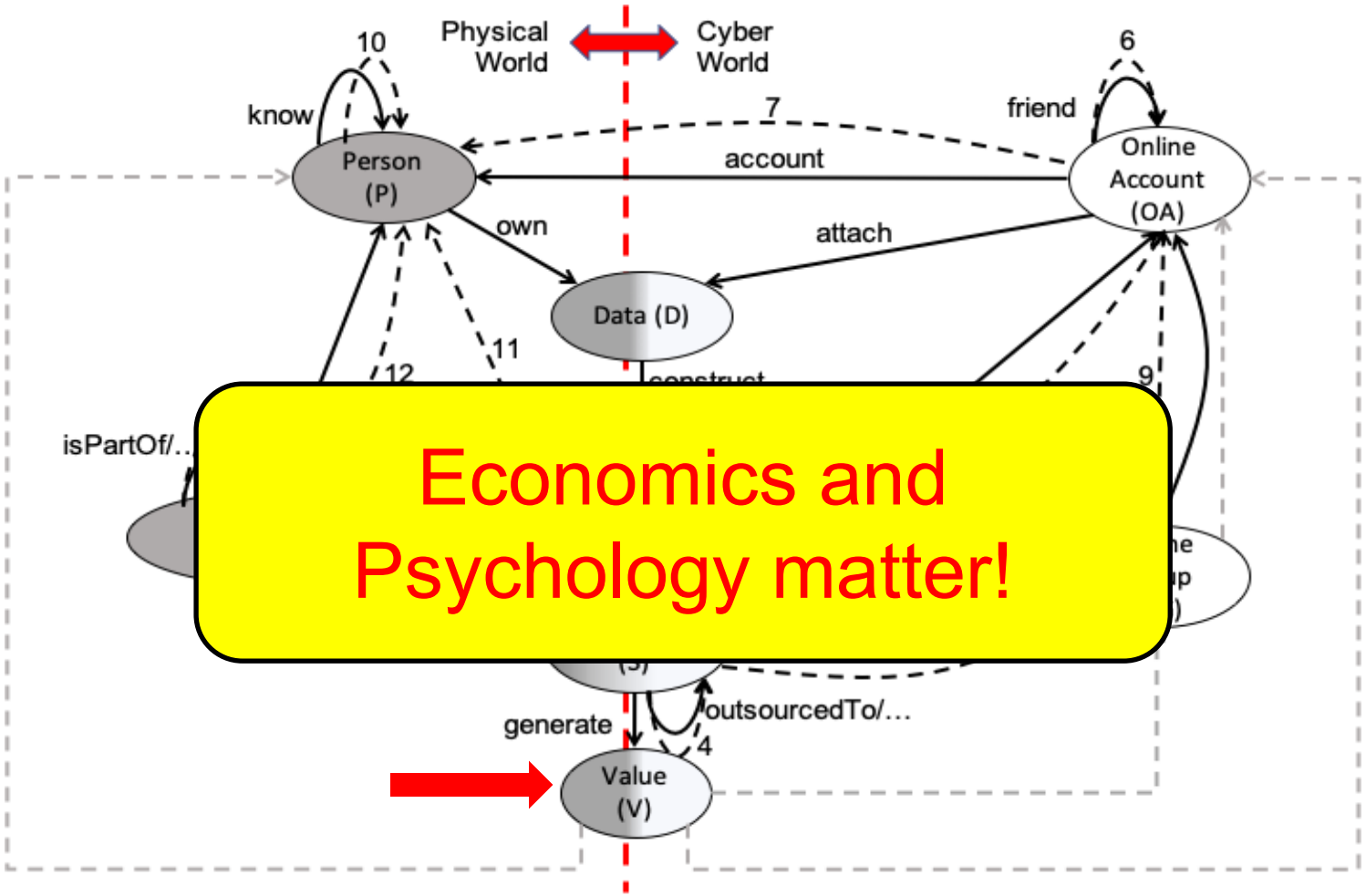
Automatic reasoning is possible!

University of Kent | Institute of Cyber Security for Society (iCSS)



Adding returned values

University of Kent | Institute of Cyber Security for Society (iCSS)



Yang Lu and Shujun Li (2022) From Data Flows to Privacy-Benefit Trade-offs: A User-Centric Semantic Model. *Security and Privacy*, 5(4):e225, 24 pages, John Wiley & Sons, Inc. doi: 10.1002/spy2.225

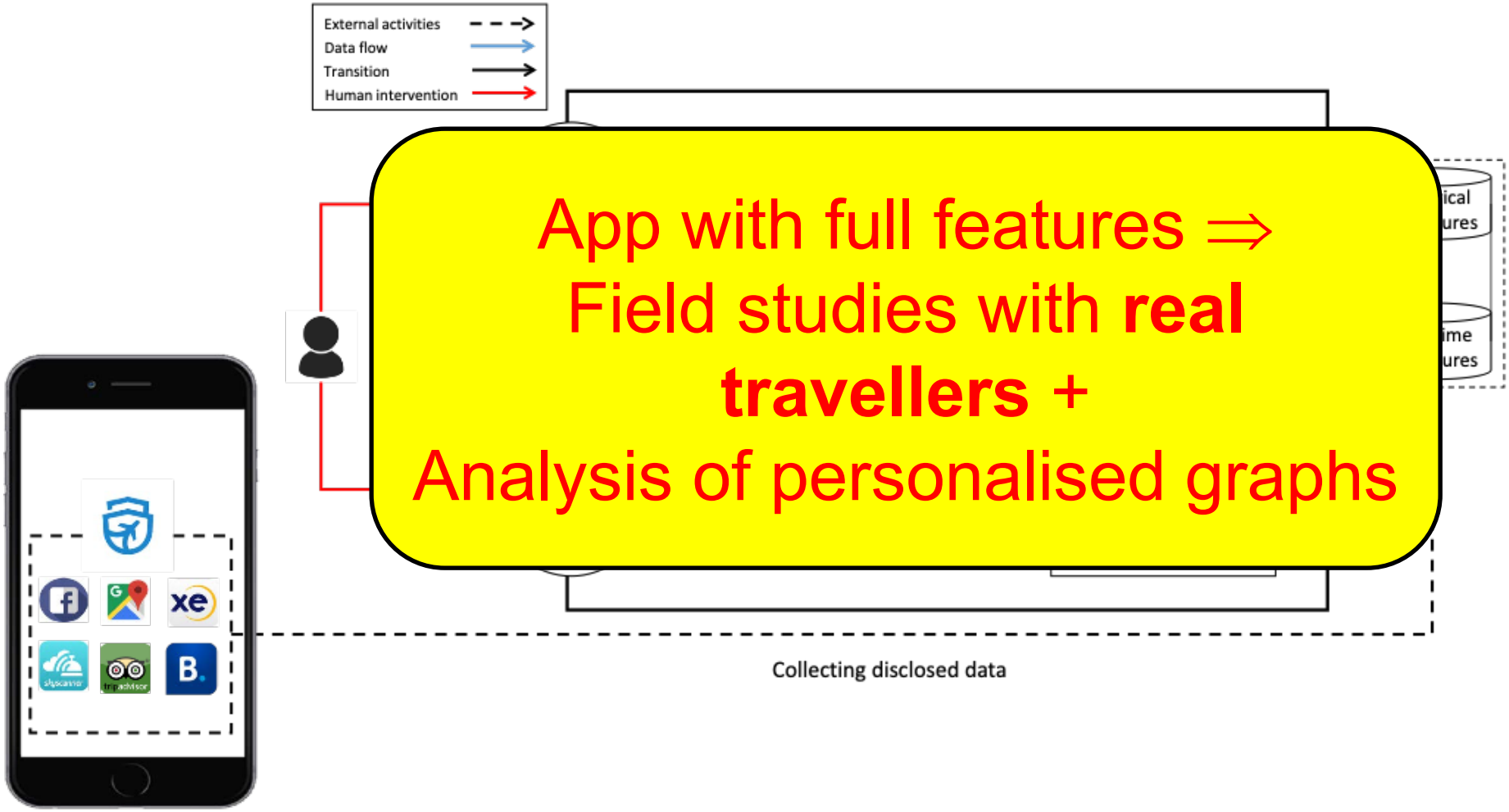
Personalized data flow graphs

University of Kent | Institute of Cyber Security for Society (iCSS)

- **User-centric** and **service-independent** tools are needed to build “my” data flow graph.
- ⇒ We are developing an Android app for this.

The vision: user-centric, server-less, from privacy awareness to nudging

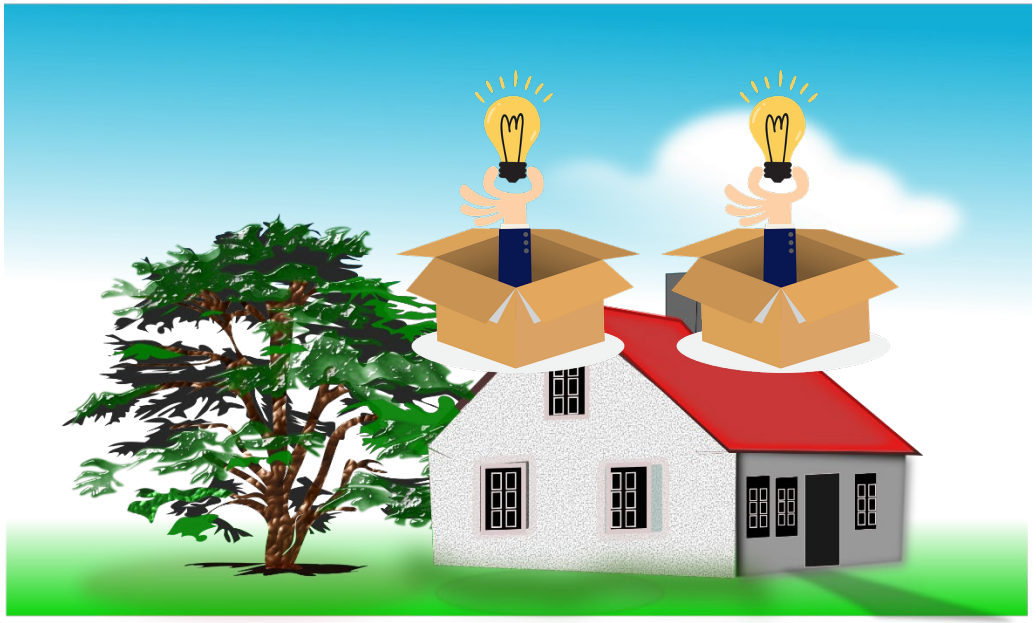
University of Kent | Institute of Cyber Security for Society (iCSS)



Yang Lu, Shujun Li, Athina Ioannou and Iis Tussyadiah (2019) From Data Disclosure to Privacy Nudges: A Privacy-aware and User-centric Personal Data Management Framework. In *Proc. DependSys 2019*, Springer. doi:10.1007/978-981-15-1304-6_21

University of Kent | Institute of Cyber Security for Society (iCSS)

Take-Home Messages



Anything useful from the talk?

University of Kent | Institute of Cyber Security for Society (iCSS)

- Data privacy problems **in all domains** can be studied using **data flow graphs**.
- Such data flow graphs are **personalised** around a node called “me”.
  - **User-centric** tools are needed to engage users.
- There are essential research questions across **multiple disciplines**.
  - There are **legal**, **ethical** and **economic** implications!
- Although PriVELT as a project has completed, our work is still **ongoing**.
  - $\Rightarrow$ We welcome different types of **collaborations**!

Privacy through lens of data flows

University of Kent | Institute of Cyber Security for Society (iCSS)

Shujun LI (李树钧)

http://www.hooklee.com/