

TILTING perspectives 2017: Regulating a connected world
Tilburg, 19 May 2017

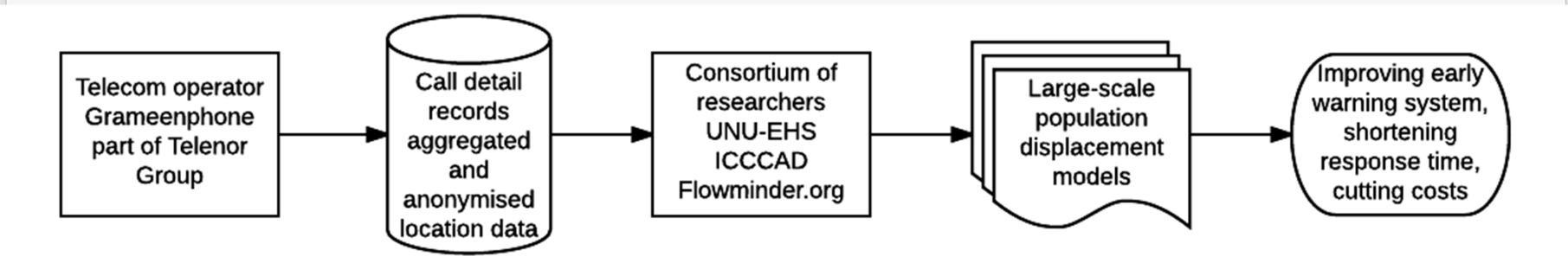


Data sharing mechanisms and privacy challenges in Data Collaboratives: Delphi study of most important issues

Ella Kolkowska (Örebro University)

Iryna Sussha (Örebro University/TU Delft)

Bastiaan van Loenen (TU Delft)



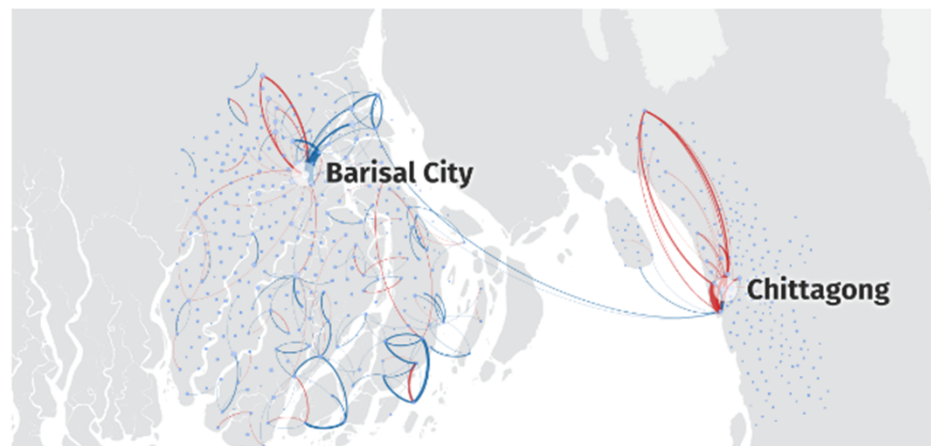


GUEST CONTRIBUTOR

Mobile Phone Data Helps Identify Displaced People Faster, Cheaper, More Accurately

August 18, 2016 | By [David J. Wrathall](#) & [Xin Lu](#)

August 18, 2016 | By *David J. Wrathall & Xin Lu*



JOIN THE CONVERSATION



Circle of Blue Retweeted



Circle of Blue
@circleofblue

Event: Water and security in an uncertain world. October 19. Webcast will be available: [@newsecuritybeat bit.ly/2dAhLBH](http://@newsecuritybeat.bit.ly/2dAhLBH)

Megan M. Roberts
@MeganMRoberts1

UN's #NextSG Brings Humanitarian Experience, But Will It Matter? from @cassidyjoseph goo.gl/5ZFIAJ @NewSecurityBeat



New UN Secretary-Ge...

In first, Uber to share ride data with Boston

Information on users will be cut



Top 10 Trending Articles

Most Viewed

Most Commented

Most Shared

Blowback for American sins in the Philippines

Trump's supporters talk rebellion, assassination at his rallies

Four women in Springfield overdose on heroin

Enough is enough — scrap the third debate

Hillary Clinton shuns spotlight as Donald Trump spirals

Vote all you want. The secret government won't

B You're reading **1 of 5** free articles. Get UNLIMITED access for only 99¢ per week. **Subscribe now** >



datacollaboratives.org

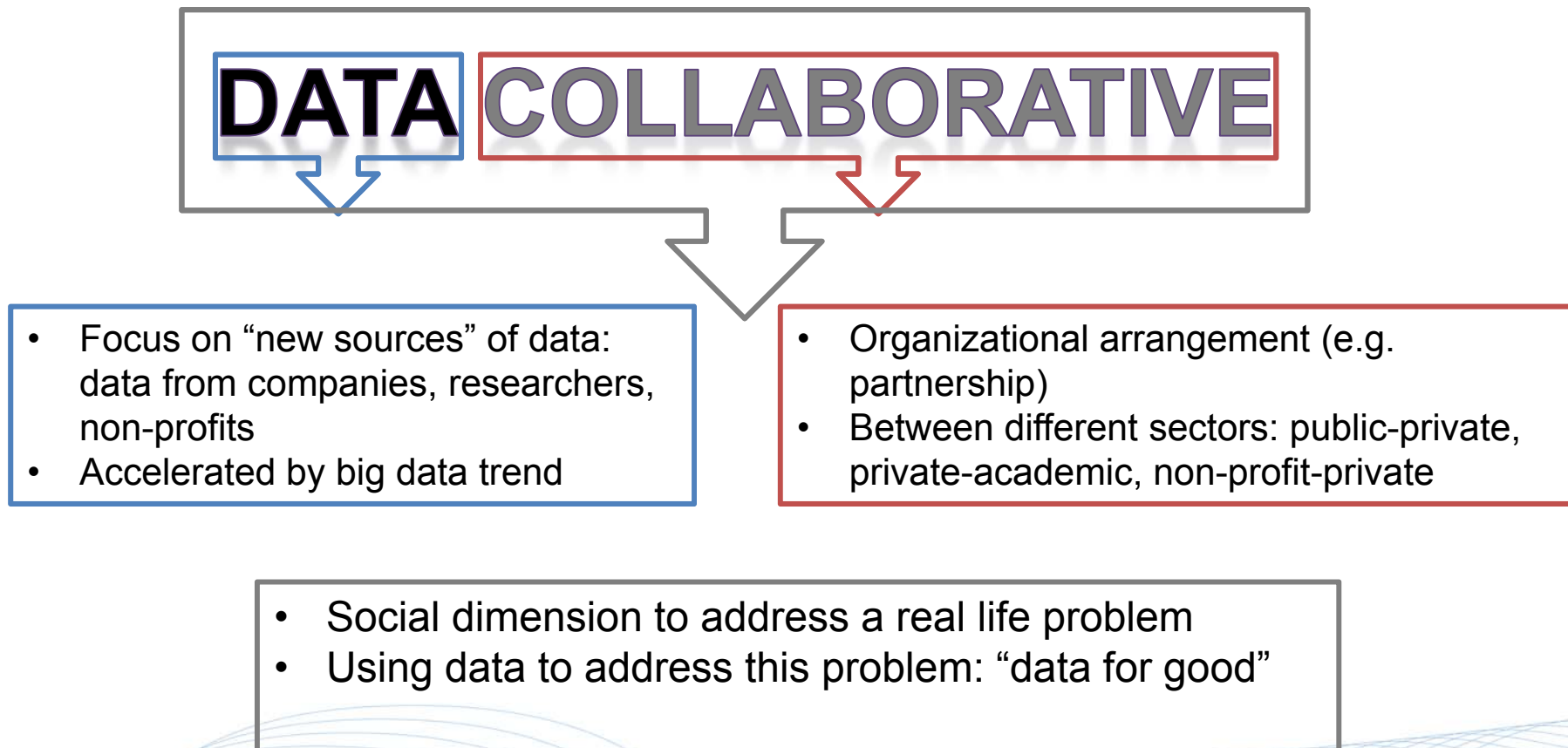
GOVLAB

DATA COLLABORATIVES EXPLORER

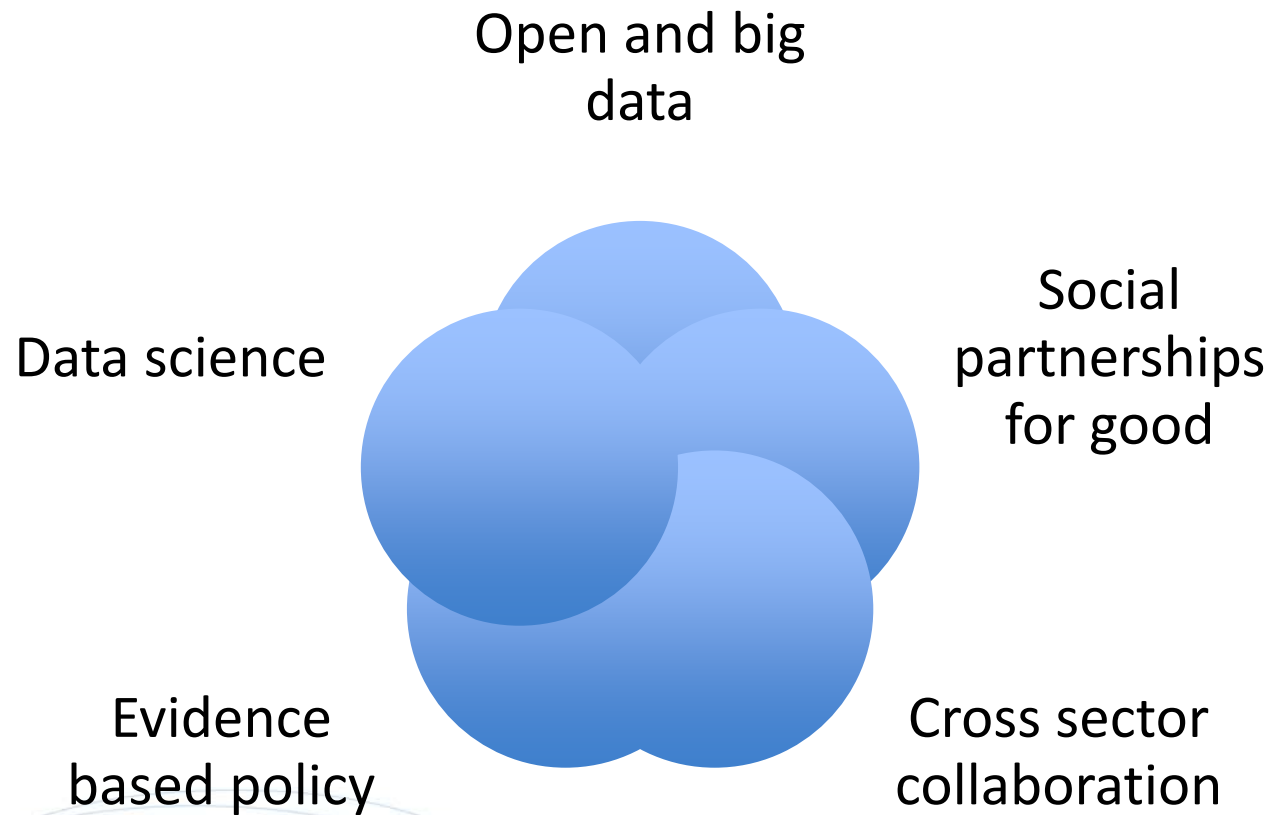
FILTERS Filter by Region Filter by Sector Filter by Collaborative Type Filter by Data Type X

COLLABORATIVE	TYPE	SECTOR	DATA TYPE	REGION
23andMe Patient-Centric Research Portal	Intelligence Product, Research Partnerships	Health	Disclosed Personal Data	Worldwide
AACR Project Genomics Evidence Neoplasia Information Exchange	Data Pooling, Trusted Intermediary, Research Partnerships	Health	Disclosed Personal Data	North America
Accelerating Medicines Partnership (AMP)	Data Pooling	Health	Disclosed Personal Data	North America
BBVA's Innova Challenge	Prizes & Challenges	Economic Development	Observed Personal Data	Europe and Central Asia

Intro: what is a data collaborative?



Intro: what is a data collaborative?



Intro: what is a data collaborative?

- Diverse domains: humanitarian and development action, healthcare, environment, poverty reduction, education etc.
- Variety of data sources: data from mobile apps, personal sensors, search engines, social networks, financial transactions etc.

Problem statement

- A lot of potential for DCs, but facing a challenge to balance the societal value derived from data with privacy of individuals
 - Recent research found that re-identification of individuals is possible without PII just from metadata (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013; de Montjoye, Radaelli, Singh, & Pentland, 2015)
- ❑ Current frameworks for data and privacy protection **do not account for big data collected by the privacy sector** (Global Pulse Privacy Advisory Group, 2016)
 - ❖ Compliance unclear when repurposing those data for societal benefit
 - ❖ Can hinder sharing of important data when there is an acute need for them
 - ❖ Can stall or lead to failure of DCs



Bloomberg L.P.

Bloomberg the Company & Its Products | Bloomberg Anywhere Remote Login | Bloomberg Terminal Demo Request

Bloomberg

Markets

Tech

Pursuits

Politics

Opinion

Businessweek

Sign In
Subscribe

Privacy Fears Over Student Data Tracking Lead to InBloom's Shutdown

The collapse of InBloom marks a backlash against the personalized learning industry

by **Olga Kharif**
May 2, 2014, 7:21 PM GMT-2





Photo illustration by 731; Photograph by Getty Images

A year ago, every public school student in New York State fell under the watchful eye of InBloom, a data analytics company. Schools sent the company an enormous batch of data spanning 400-odd fields that included a wide range of personal details, from test scores and special-education enrollment to whether kids got free lunches. The idea was to compile enough information so teachers or software could tailor assignments to each student's needs. InBloom had contracts to do the same for millions of public school kids

LIVE TV

AUDIO

Our objective

- *To identify measures for enhancing privacy protection in the context of Data Collaboratives*
- Scope:
 - ❖ Focus on DCs which involve the sharing of data about private persons
 - ❖ Focus on DCs which are aimed at sharing by companies of already collected data

Literature review

- Literature specifically at the intersection of data collaboratives and privacy is limited
- However research discussing privacy challenges and implications in relation to big data
- Most of these challenges either discuss privacy in relation to big data in general or in relation to a certain type of data
 - call detail records

	Privacy challenge	Summary	References
1	Lack of informed consent	Consent when using services VS informed consent for human subjects research	Kahn et al. (2014); Wyber et al. (2015); Taylor (2016)
2	Risk of re-identification of persons	Removing PII makes re-identification only slightly more difficult	de Montjoye et al. (2014); Hoepner et al.
3	Ambiguous and country-specific legislation	Frameworks do not specifically address big data which leads to self-regulation by companies	de Montjoye et al. (2014); UN Global Pulse (2016); Taylor & Schroeder (2015); Wyber et al. (2015), Latonero & Gold (2015)
4	Group privacy	Even de-identified data may be sensitive and may allow people to be tracked as groups and networks	Taylor 2016
5	Justification of re-identification of persons at risk	Re-identification may be done to contact the person if they are at risk	(Latonero & Gold, 2015)

Method: ranking type Delphi (Schmidt, 1997)

1st round
Discovery of
issues

- Respondents submit as many issues as possible
- Researchers create consolidated list of all issues
- Respondents verify the consolidated list

2nd round
Determining of
most important
issues

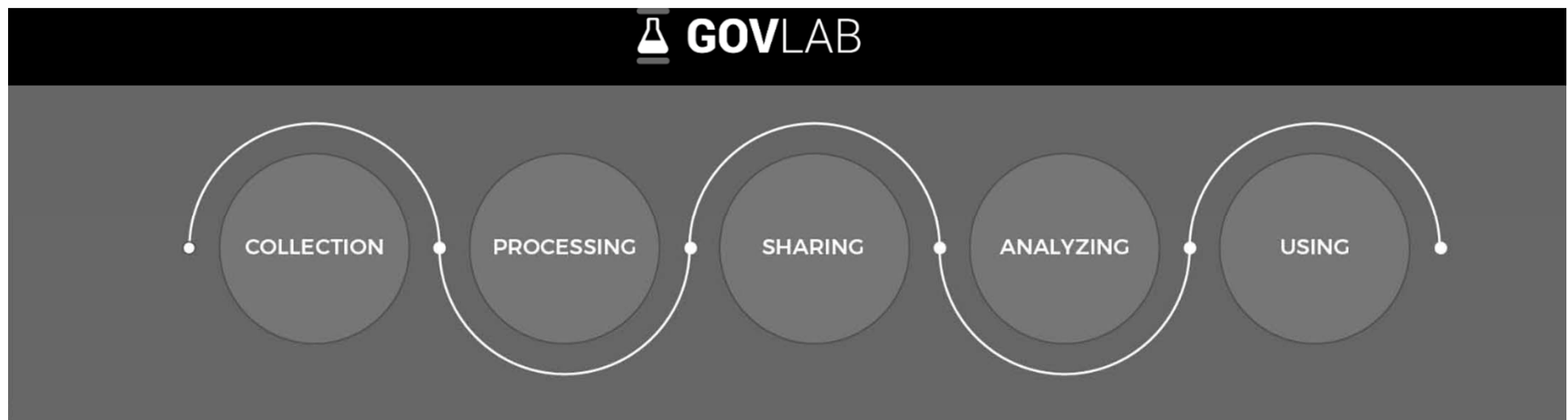
3rd round
Ranking the
issues

Results

- 9 experts participated in the pilot 1st round
- Academics and practitioners on data collaboratives (3), privacy (4), and data analytics (2)
- Survey conducted during March-April
- 63 measures to enhance privacy identified

Analysis matrix

- Dimension of privacy it concerns → inductively formulated in iterative way
 - Legal, Norms, Organisation, Procedures, Technology ...
- Stage of the data collaborative data lifecycle it concerns



DIMENSIONS OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			
TECHNICAL	Use of a privacy dashboard for data subjects to control their data (21)				
			End users work on an institutional server (60)		

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			
TECHNICAL	Use of a privacy dashboard for data subjects to control their data (21)				
			End users work on an institutional server (60)		
ORGANISAT IONAL	Appointment of a Data Protection Officer (12)				
	Adversarial testing: an independent red-team exercise to empirically ascertain the risk of privacy being violated and the effectiveness of controls. (63)				

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			
TECHNICAL	Use of a privacy dashboard for data subjects to control their data (21)				
			End users work on an institutional server (60)		
ORGANISAT IONAL	Appointment of a Data Protection Officer (12)				
	Adversarial testing: an independent red-team exercise to empirically ascertain the risk of privacy being violated and the effectiveness of controls. (63)				
POLICY		Written agreement (MOU) between data provider and data users (7,9,18, 25b)			
		Data access control policies (2a,24)			

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			
TECHNICAL	Use of a privacy dashboard for data subjects to control their data (21)				
			End users work on an institutional server (60)		
ORGANISAT IONAL	Appointment of a Data Protection Officer (12)				
	Adversarial testing: an independent red-team exercise to empirically ascertain the risk of privacy being violated and the effectiveness of controls. (63)				
POLICY		Written agreement (MOU) between data provider and data users (7,9,18, 25b)			
		Data access control policies (2a,24)			
PROCEDUR ES		Auditing and accountability of access policies (2b, 24)			
	Opt in or opt out available to data subjects at all stages (35, 57c).				

DIMENSION S OF PRIVACY MEASURES	DATA COLLABORATIVE DATA LIFECYCLE STAGES				
	COLLECTING	PROCESSING	SHARING	ANALYSING	USING
LEGAL		Oversight by data protection authorities (25a)			
		Self-regulation, industry standards and codes of practice (32, 51)			
NORMS				Justified requirements for the use of data (45)	
		Privacy by design (6)			
TECHNICAL	Use of a privacy dashboard for data subjects to control their data (21)				
			End users work on an institutional server (60)		
ORGANISAT IONAL	Appointment of a Data Protection Officer (12)				
	Adversarial testing: an independent red-team exercise to empirically ascertain the risk of privacy being violated and the effectiveness of controls. (63)				
POLICY		Written agreement (MOU) between data provider and data users (7,9,18, 25b)			
		Data access control policies (2a,24)			
PROCEDUR ES		Auditing and accountability of access policies (2b, 24)			
	Opt in or opt out available to data subjects at all stages (35, 57c).				

- Most “populated” categories: Technical, Policy, Procedures

Linking measures to challenges from the literature



Privacy challenge	Survey response/ measure suggested
1 Lack of informed consent	<ul style="list-style-type: none">• Use of a privacy dashboard for data subjects to control their data (21)• Company's privacy policy should notify the user when data sharing in a data collaborative might arise (34)
2 Risk of re-identification of persons	<ul style="list-style-type: none">• Adversarial testing: independent red-team exercise to empirically ascertain the risk of privacy being violated and the effectiveness of controls (63)• Linking restraints: end users could be restrained from linking datasets in a manner which might reveal individuals (56)
3 Ambiguous and country-specific legislation	<ul style="list-style-type: none">• Self-regulation, industry standards and codes of practice (32, 51)
4 Group privacy	Not mentioned
5 Justification of re-identification of persons at risk	<ul style="list-style-type: none">• Justified use: end users should justify the requirements for using data (45)• DPIA (Data Protection Impact assessment) (54)

- Many measures relate to all challenges as a whole: e.g.
 - Written agreement (MOU) between data provider and data users
 - Implementing a privacy governance structure
 - Etc.

Our next steps

- Validate our matrix with the respondents and ask if they can think of any additional measures
- Move on with the 2nd phase of the study – determining of most important measures

Questions for discussion

- Are any significant elements missing from our framework/findings? Any unexpected results?
- Any suggestions of an existing framework of privacy dimensions which we can use?
- Any thoughts on the prospects of our study? Any suggestions for next stages?



Contact: iryna.susha@oru.se