At Crossroads Between

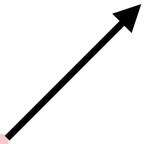
Personalisation and Privacy

Harshvardhan J. Pandit | pandith@tcd.ie | @coolharsh55

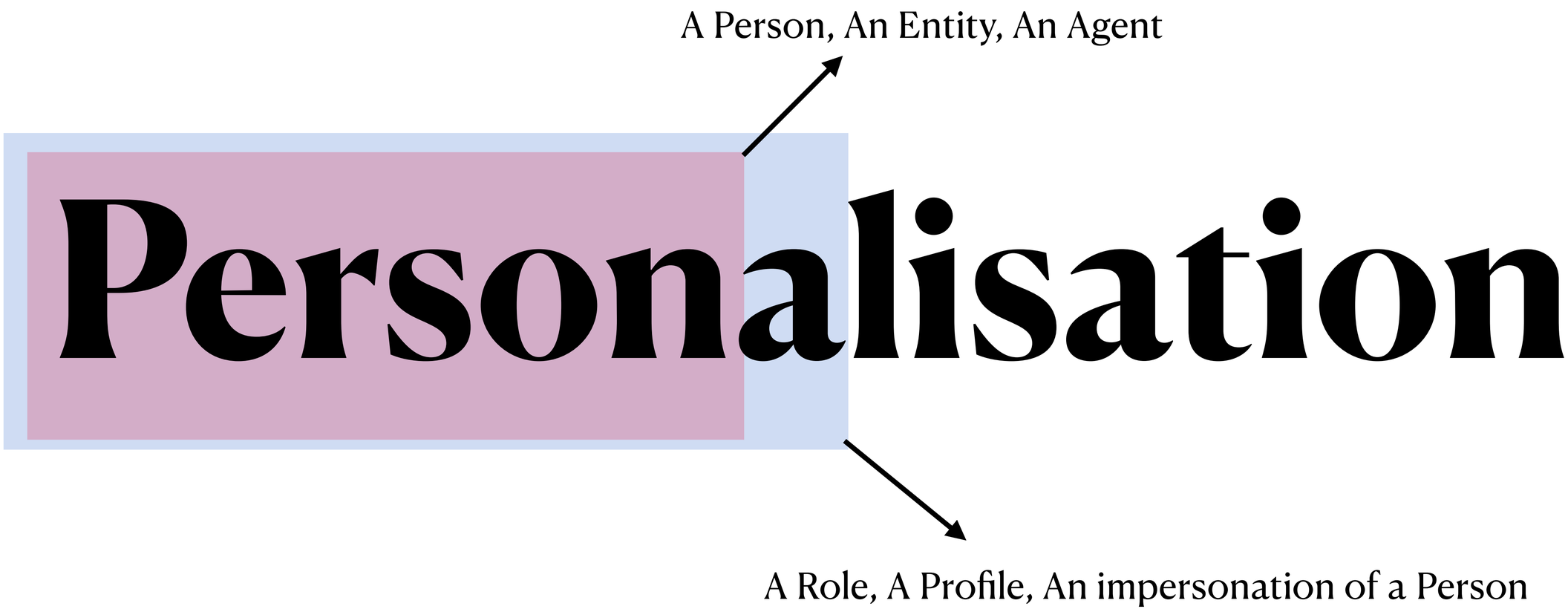
ADAPT PEA Reading Group | 03 March 2022 | Dublin City University

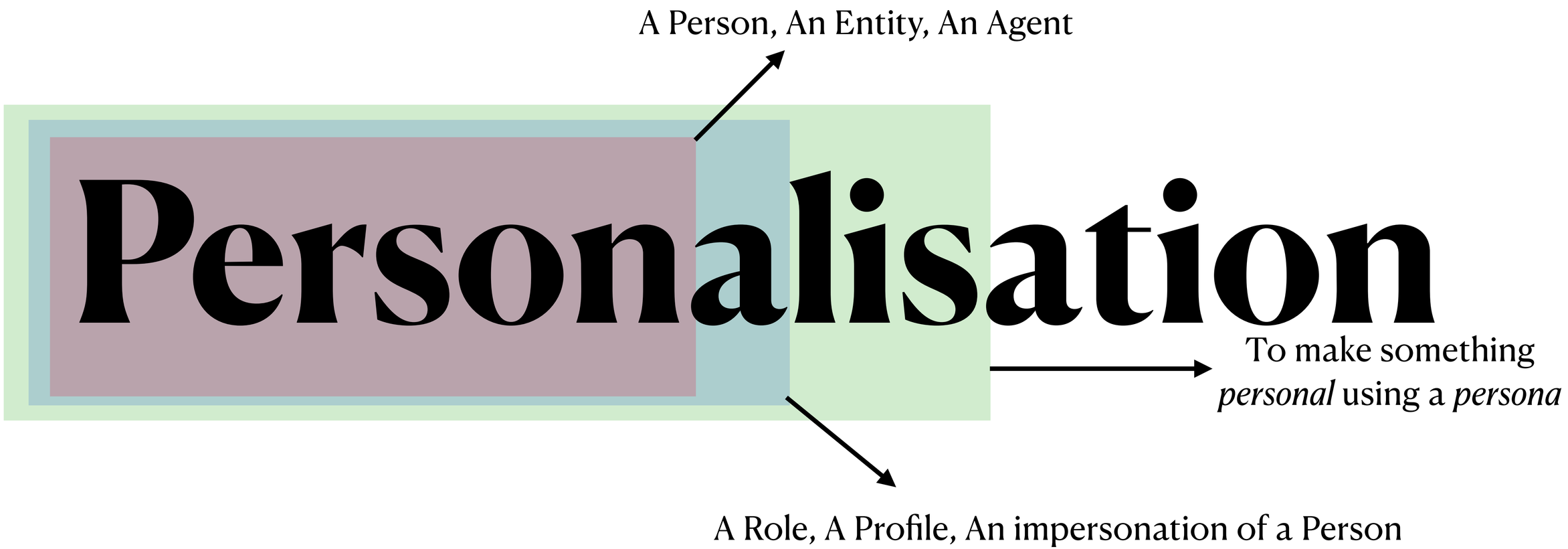
Slides available at: https://harshp.com/research/presentations

A Person, An Entity, An Agent



Personalisation


A Person, An Entity, An Agent
Personalisation
A Role, A Profile, An impersonation of a Person


A Person, An Entity, An Agent
Personalisation
To make something
personal using a *persona*
A Role, A Profile, An impersonation of a Person

Privacy

Seclusion, Private, Secrecy, Concealment, Knowledge, Permission, Control, Sensitivity, Anonymity

GDPR et al.

Personalisation needs Personal Data ; Personal Data is regulated by GDPR

- Personal Data :: Sensitivity, “Special Category”, PII
- Legal Basis :: Contract (e.g. Provide a Service -> Netflix Recommendations)
- Legal Basis :: Legitimate Interest (e.g. personalised demographic ads)
- Legal Basis :: Consent (e.g. ask to personalise Ads on websites)
- Principles :: Data Minimisation (use only what is needed)
- Data Protection Impact Assessment :: Any potential impacts? Harms?

Personalisation vs Privacy

Availability of Information Reduces Privacy but Increases potential for Personalisation

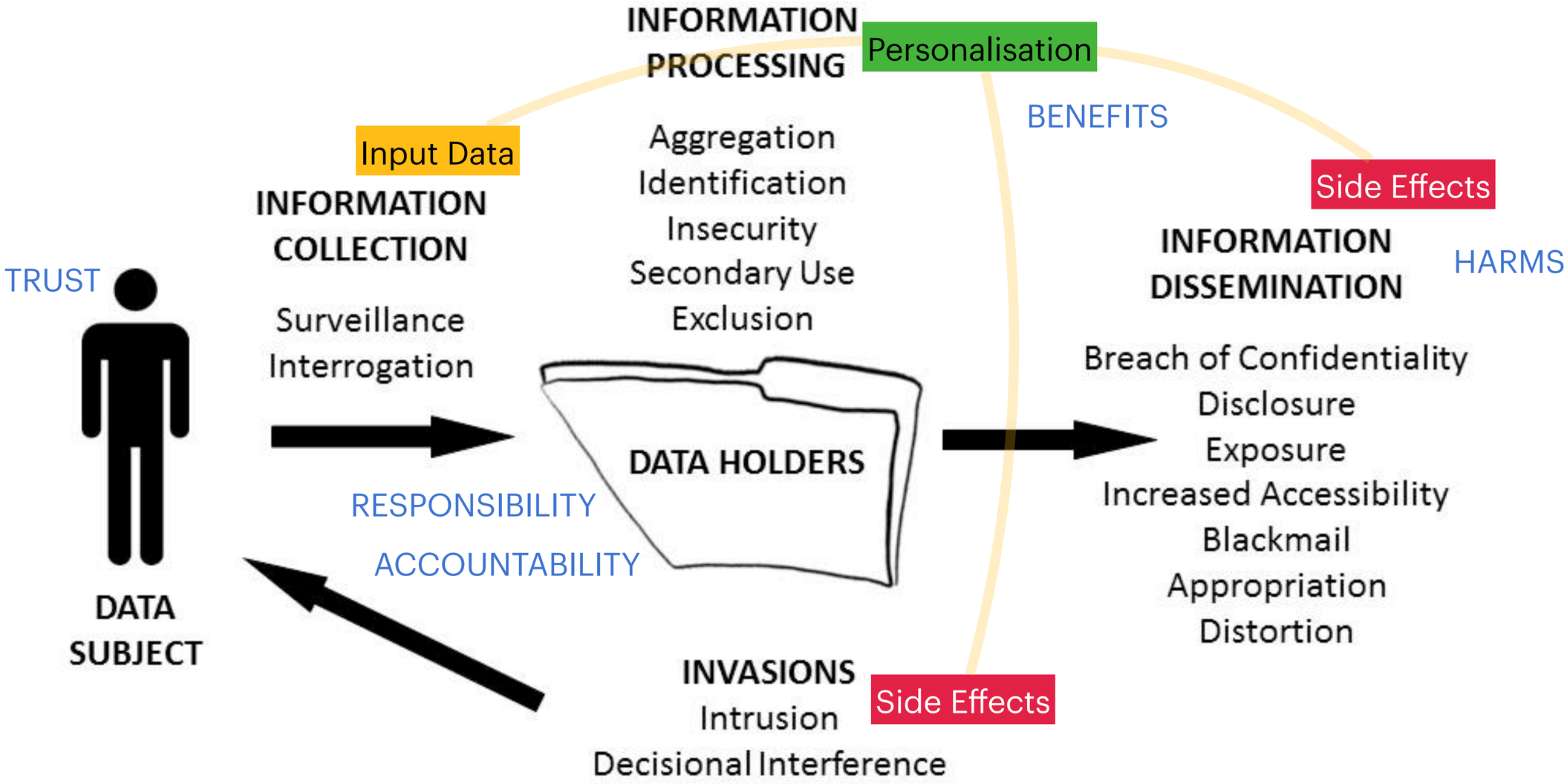
Privacy as Confidentiality
- Data Minimisation
- Reduce data required
- Identifiability of 'Persons'
- Necessity of Data Required

SECURITY

Privacy as Control
- Data Protection, Notices
- Control Flows, Algorithmic
- Involvement, Agency, Ability
- Rights, Norms

Privacy as Practice
- Contextual, Mediation
- Feedback, Negotiation
- Dynamic, Interactive
- Norms, Guidelines

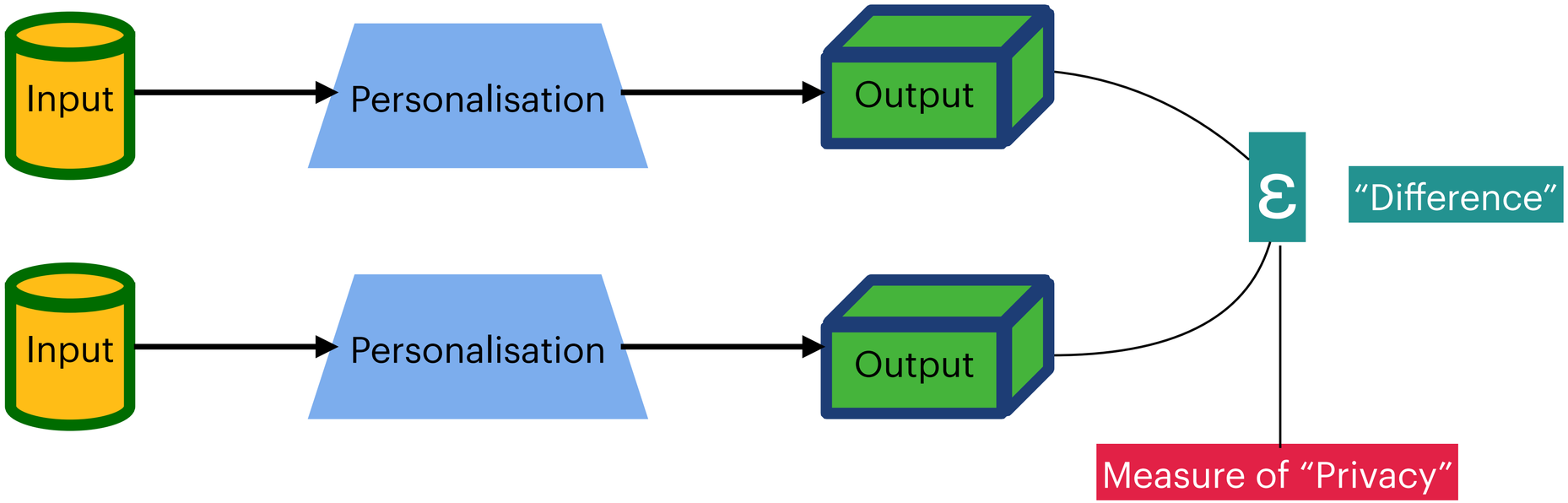
Can you engineer privacy? On the potentials and challenges of applying privacy research in engineering practice - Seda Gurses
https://www.esat.kuleuven.be/cosic/publications/article-2465.pdf



Taxonomy of Privacy - Daniel Solve https://ssrn.com/abstract=667622

Differential Privacy

Performing Personalisation with lesser loss of Privacy



Differential Privacy: A Primer for a Non-Technical Audience - Wood et al. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3338027

Differential Privacy

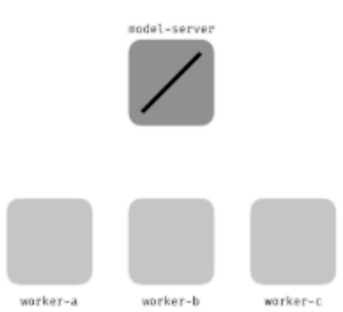
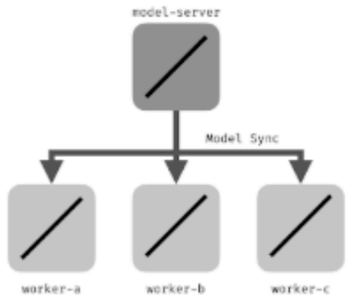
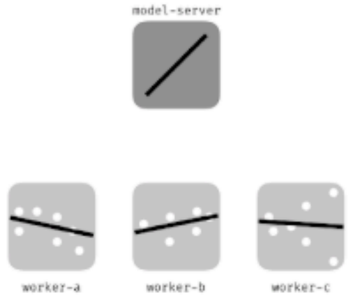
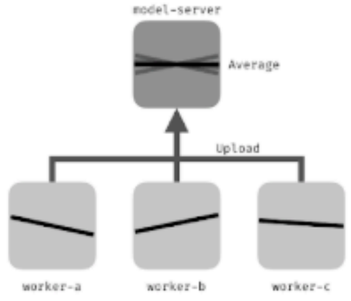
Statistical Measurements and Guarantees

- Privacy ≋ Access to Data / Identifiability
- Introduce ‘randomness’ to outputs to protect ‘privacy’
- Calculate ‘Risk’ of ‘Privacy Loss’
- Create a ‘Privacy Budget’
- Guarantees regarding Accuracy and Performance
- Group Privacy

Differential Privacy: A Primer for a Non-Technical Audience - Wood et al. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3338027

Federated Learning

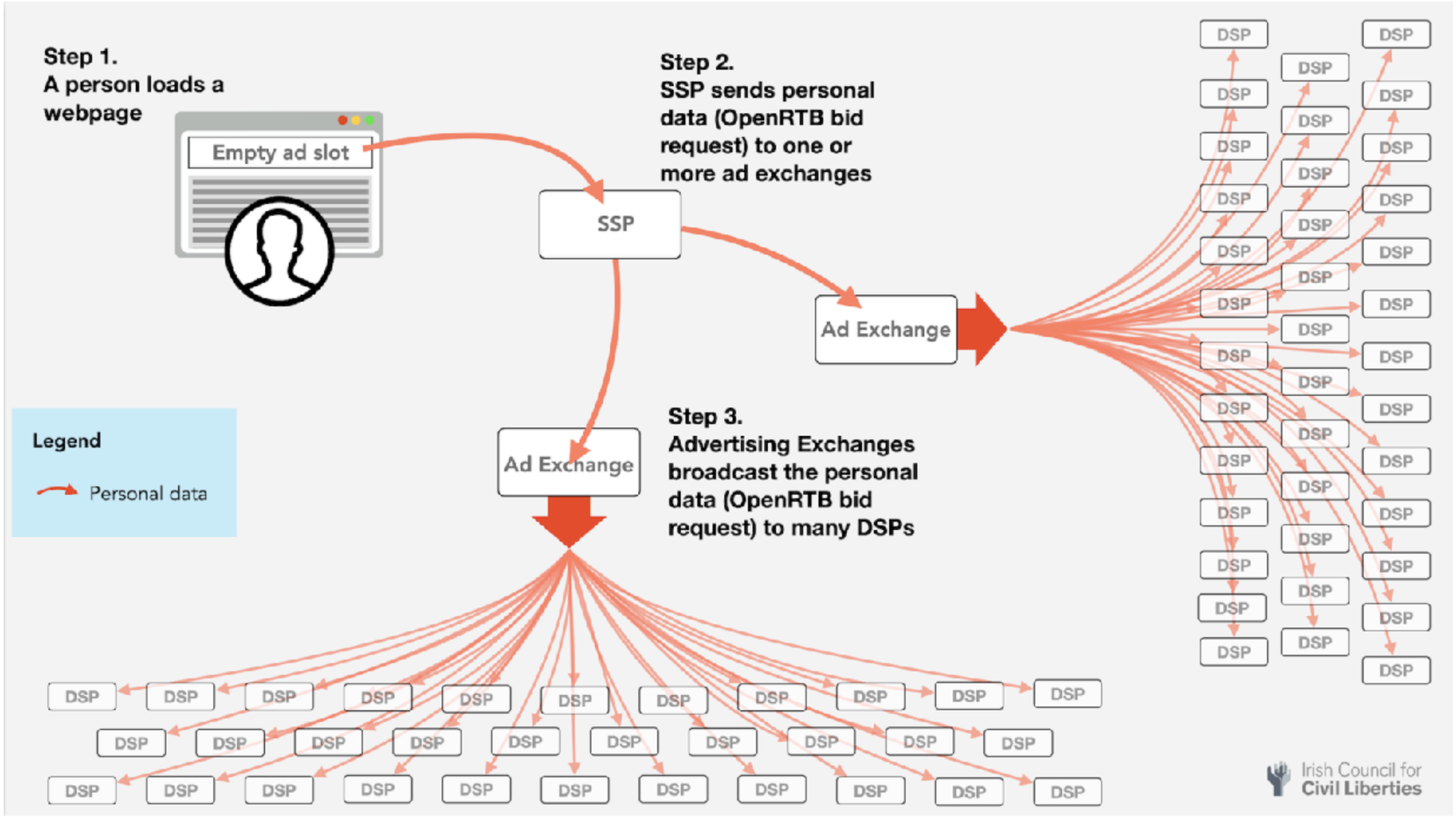
Do ML locally and pool models globally

Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

https://en.wikipedia.org/wiki/Federated_learning

Current Personalised Advertising Model

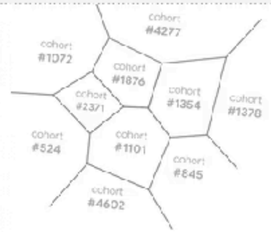
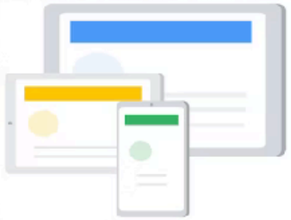
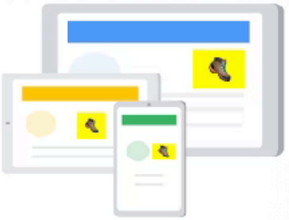
Surveillance-based Targeted Advertising



https://www.iccl.ie/digital-data/iab-europe-cant-audit-what-1000-companies-that-use-its-tcf-system-do-with-our-personal-data/

Google's FLoC Proposal

Federated Learning of Cohorts uses 'cohorts' to target advertisements

FLoC service	Browser	Advertiser *a site that pays to advertise its products* shoestore.example	Publisher *a site that gets paid to display ads* dailynews.example	Adtech *a company that provides services to deliver ads* adnetwork.example	Publisher *a site that gets paid to display ads* dailynews.example
					
Create FLoC model	**1 Calculate browser cohort**	**2 Observe cohort activity**	**3 Observe visitor cohorts**	**4 Select ads relevant to cohort**	**5 Display relevant ads**
The FLoC service creates a mathematical model with thousands of "cohorts". Each cohort corresponds to thousands of web browsers with similar recent browsing histories.	A user's browser, on their device, uses the FLoC model to calculate which cohort corresponds most closely to its own browsing history.	An advertiser observes the activity of cohorts on its site and shares this with its adtech platforms.*	A publisher observes the cohorts of visitors to its site and shares this with its adtech platforms.*	An adtech platform selects ads appropriate for the user's cohort.	The publisher displays ads that are relevant to the user based on their browser's cohort.

* The adtech platform may handle observing the cohort information on behalf of the advertiser or publisher.

https://developer.chrome.com/docs/privacy-sandbox/floc/

Overview of Personalisation Issues

Key takeaways

- What data is ‘used’ ??? —> Transparency
- What data is ‘needed’? What is ‘necessary’? —> Data Minimisation
- What are the sources of ‘data’ ? —> Transparency
- Is any data ‘sensitive’ ? Is it ‘special’ ? —> Ethical Concerns
- Is data (input/output) ‘accurate’ —> Accountability
- Is the output configurable ? —> Privacy by Design / Default
- Understand distinctions between *Privacy* vs *Security* vs *Identifiability* vs *Control*

At Crossroads Between

Personalisation and Privacy

Harshvardhan J. Pandit | pandith@tcd.ie | @coolharsh55

ADAPT PEA Reading Group | 03 March 2022 | Dublin City University

Slides available at: https://harshp.com/research/presentations