

# Workshop #4

# Innovative Models and Data Management

## Session 1

## Workshop #4 Session 1

Some information for a smooth running of the session.



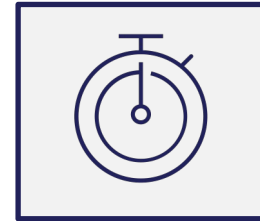
The audio of the session is recorded.  
**Please only speak in the microphone.**



**Please introduce yourself every time before speaking.**  
Do not hesitate to participate and ask questions.



Avoid using your smartphones and laptops if possible.



**Please make sure to respect the time allocated** for your speech & keep your interventions under 1mn.  
Please be on time to the next session.

## Workshop #4 Session 1

### Facilitators

**Jaime GUIDRY AUVIL** - US National Cancer Institute

**Gijs GELEIJNSE** - Netherlands Comprehensive Cancer Organisation

### Subthemes

- Will anonymization provide a potential solution to data sharing for all or many countries?
- What are the potential drawbacks of anonymization as a solution?

## Discussants



**Steven THOMAS**

Canada

**Statistics Canada**

Section chief



**Leucio Antonio CUTILLO**

Italy

**European Commission,  
Joint Research Centre (JRC)**

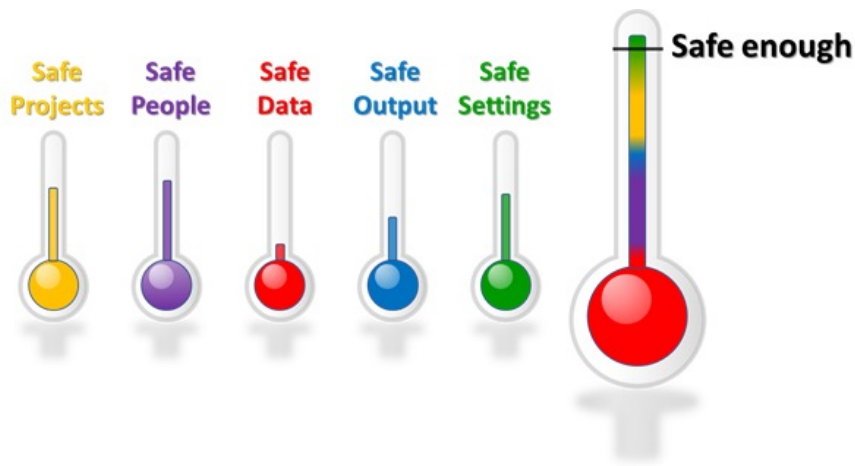
Project officer and Scientific  
researcher / Disease Prevention  
Unit

# Creating Safe Data to Facilitate International Data Collaboration

**Steven Thomas**  
**Statistics Canada**

## Access Framework – The 5 Safes (www.fivesafes.org)

- 5 Safes: Key sources of disclosure when thinking about access.
- Work toward Safe enough by aiming to achieve:
  - **Safe Projects** and/or
  - **Safe People** and/or
  - **Safe Settings** and/or
  - **Safe Outputs** and/or
  - **Safe Data**



<https://www.youtube.com/watch?v=MIn9T52mwj0>

# Safe Data

Various degrees of 'Safety' based on the degree of risk and harm

## Risk – Identifiability

- **Serious risk** - Identifiable data. Little or no effort required. Unique identifiers present (name, address, social insurance numbers).
- **Medium Risk** - Pseudonymized data. Some effort required. Direct identifiers removed. Variables that in combination can identify an individual (Indirect Identifiers - gender, age, postal code) still present.
- **Low risk** - Anonymized data. Significant effort required. Direct identifiers removed. Indirect identifiers treated (removed, adjusted). No serious possibility of reidentification.

## Harm – Severity of a breach / impact on the individual

- **Negligible** (not cause any harm) → **Severe** (Serious harm to the individual)

# Statistics Canada's Confidentiality Classification Tool

Sensitivity (Impact scale)	5	0	5	10	15	20	25	30	35	40	45	50
	4	0	4	8	12	16	20	24	28	32	36	40
	3	0	3	6	9	12	15	18	21	24	27	30
	2	0	2	4	6	8	10	12	14	16	18	20
	1	0	1	2	3	4	5	6	7	8	9	10
		0	1	2	3	4	5	6	7	8	9	10
Potential Disclosure Risk												

**Evaluation for this Product:**

Categorization of Potential Disclosure Risk

5

Categorization of Product Sensitivity

1

Combined Risk-Sensitivity Score

5

Overall Classification of Product

Level:

2

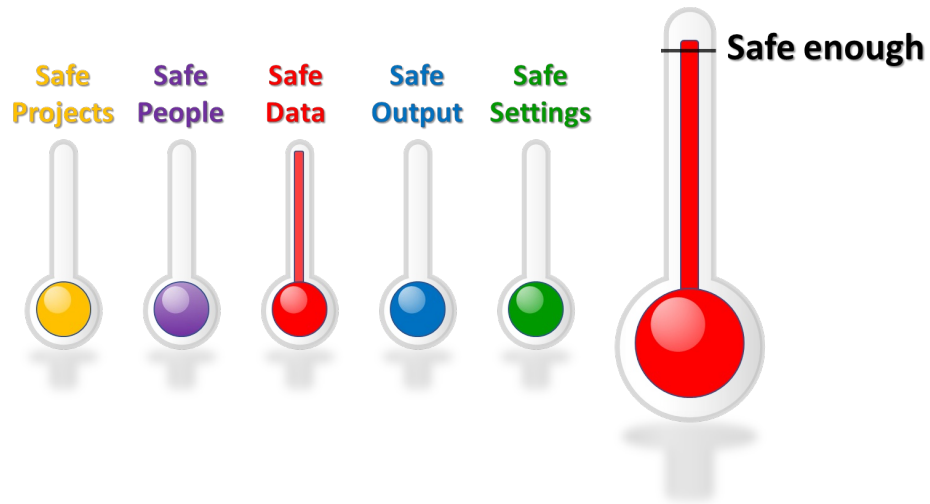
Overall Classification Levels based on Risk-Sensitivity Score:

LB	UB	Level
45	50	9
36	44	8
30	35	7
21	29	6
15	20	5
10	14	4
7	9	3
4	6	2
1	3	1
0	0	0



## International Access Options: Public Use Microdata Files (Anonymized Data)

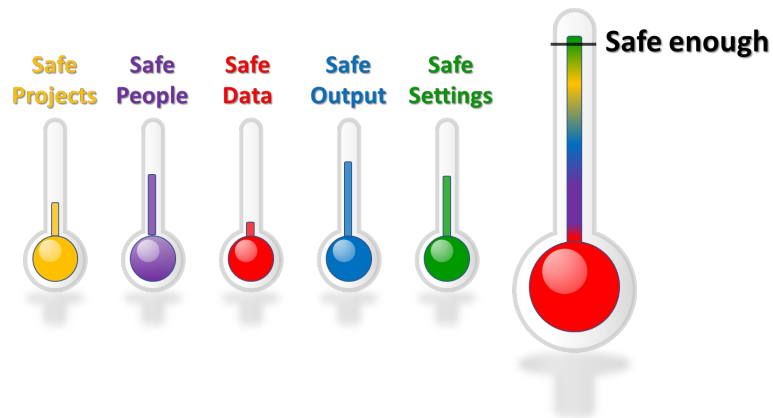
- **Safe People:** Open license agreement
- **Safe Data:** Fully anonymized data
- **Safe Setting:** Data is open
- **Safe Projects:** Any project is acceptable
  - \* Ideas can be expanded to include synthetic data
- **Safe Output:** No controls



See <https://www150.statcan.gc.ca/n1/en/type/data?MM=1> for Statistics Canada open data.

## Option 2: Privacy Preserving Access – Statistics Canada's Real-Time Remote Access<sup>1</sup> (RTRA)

- **Safe People:** Accredited users
- **Safe Data:** Basic anonymized data
- **Safe Setting:** Data remains on-site
- **Safe Projects:** Pre-approved projects
- **Safe Output:** Automated confidentiality vetting



<sup>1</sup>See <https://www150.statcan.gc.ca/eng/rtra/rtra> for more information

**Thank you! Merci!**  
**Steven Thomas**  
**Statistics Canada / Statistique Canada**  
**[Steven.Thomas@statcan.gc.ca](mailto:Steven.Thomas@statcan.gc.ca)**



# Pseudonym Generation, Pseudonym Linkage and Encrypted Pseudonymised Data Transfer with SPIDER

European Platform on Rare Disease Registration -  
European Rare Disease Registry Infrastructure

*L.A. Cutillo, Ph.D. – Joint Research Centre, Unit F.1.*

# Research area

## Problem

Pseudonym	Blood test
1234 (Alice)	BloodTest <sub>A</sub>
5678 (Bob)	BloodTest <sub>B</sub>



Hospital 1



Name	ECG
3456 (Alice)	ECG <sub>A</sub>
7890 (Carl)	ECG <sub>C</sub>



Hospital 2



Pseudonym	Blood test	ECG
???? (Alice)	BloodTest <sub>A</sub>	ECG <sub>A</sub>
???? (Bob)	BloodTest <sub>B</sub>	
???? (Carl)		ECG <sub>C</sub>



*Fragmentation* of information

*Lookup, link* and *transfer*

# Design principles



---

## Data Encrypted at the Source

Service Provider cannot decrypt data



---

## Decentralisation

No omniscient entity



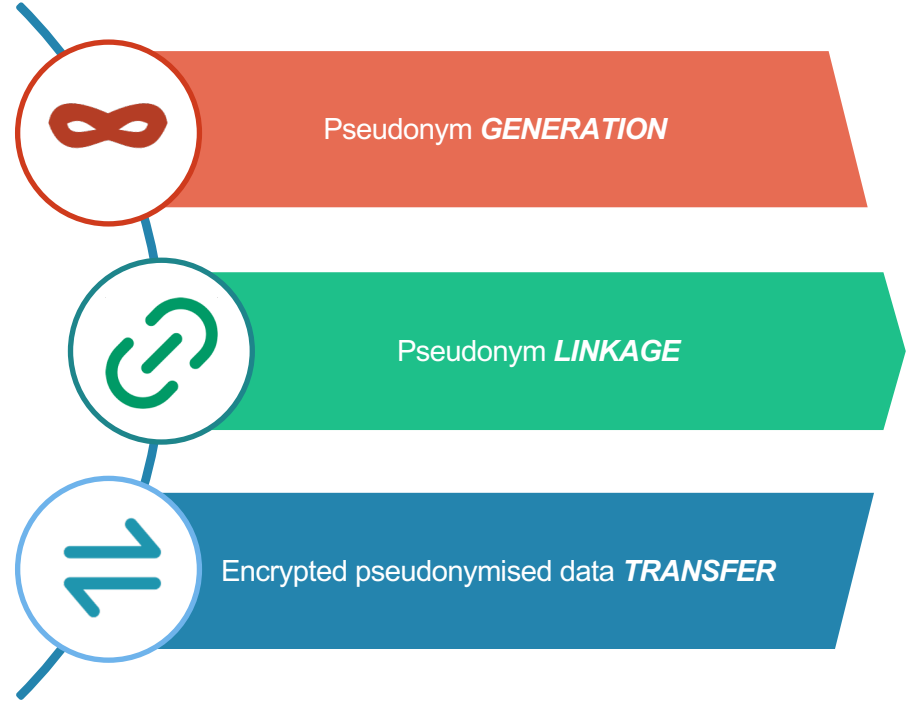
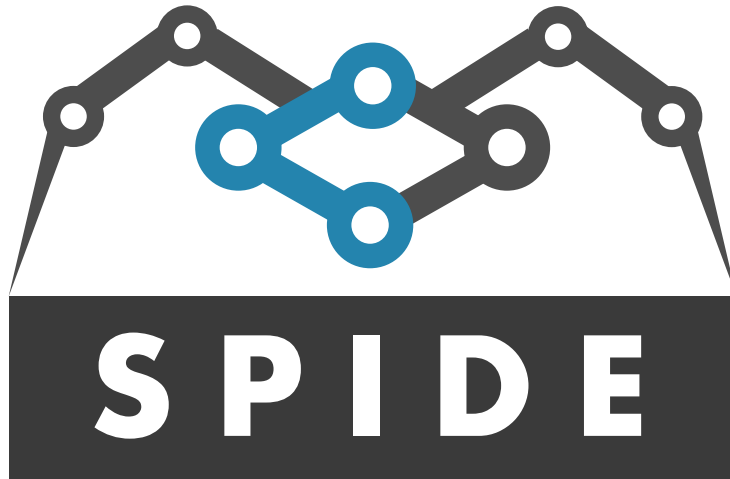
---

## Multi-layer pseudonymisation

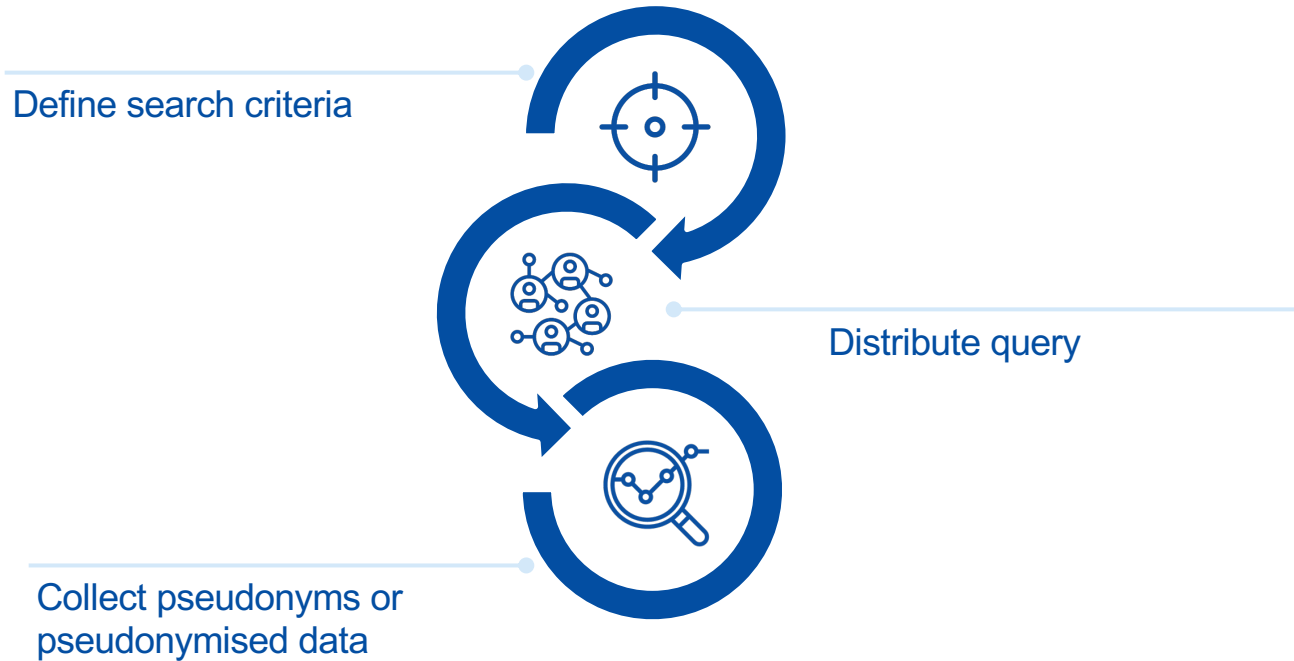
No third party re-identification

# Main Functionali

Does NOT have access to patient data!

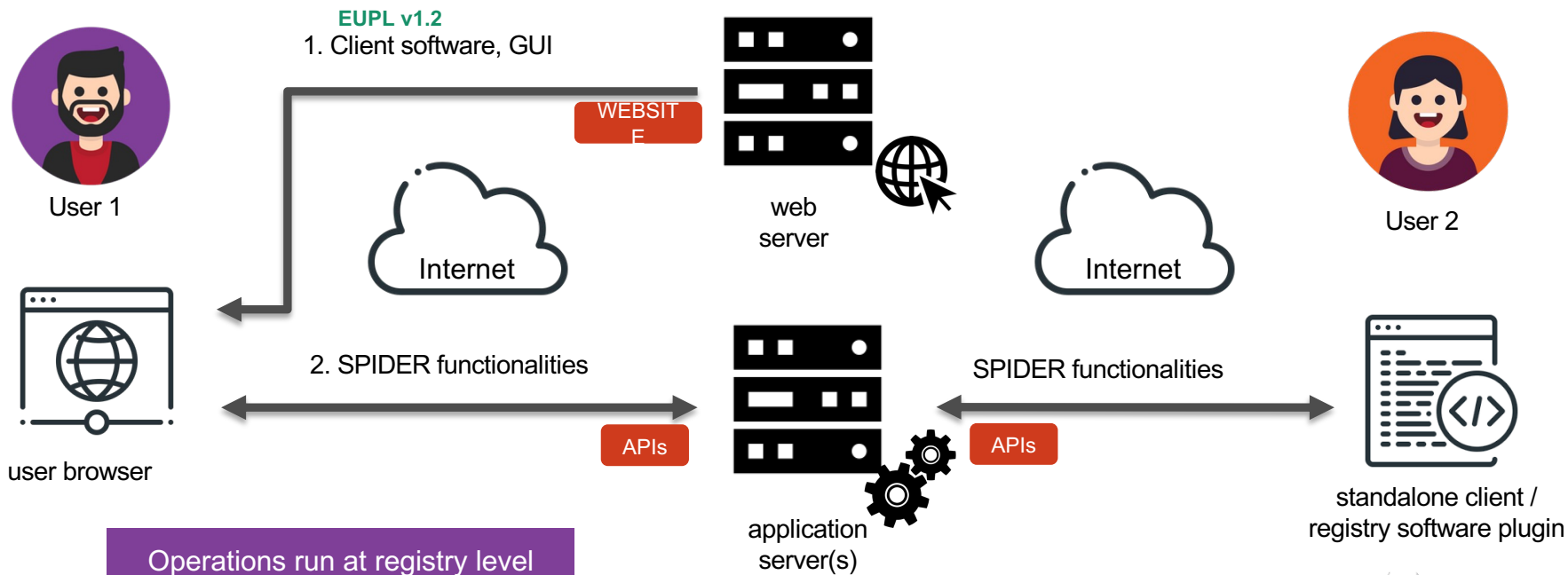


# Cohort building





# Accessing SPIDER services



# Application in the context of RD Registries

European Commission > EU Science Hub > European Platform on Rare Disease Registration > ERDRI > SPIDER

## SPIDER pseudonymisation tool

Dashboard Patients Patient groups Cohorts Notifications Registry 2

### My patients

[Add patient](#) [Search patients](#)

<input type="checkbox"/>	First name	Last name	Date of birth	Pseudonym	Additional data sources	<a href="#">view</a>
<input type="checkbox"/>	Alice	Red	01-01-2000	rgCJ	Registry 3, Registry 1	<a href="#">view</a>
<input type="checkbox"/>	Bob	Blue	02-01-2000	49lv	No additional data source for this patient	<a href="#">view</a>
<input type="checkbox"/>	Charlotte	White	03-01-2000	GTFs	No additional data source for this patient	<a href="#">view</a>
<input type="checkbox"/>	Dylan	Black	04-01-2000	Ms83	No additional data source for this patient	<a href="#">view</a>



<https://eu-rd-platform.jrc.ec.europa.eu/spider/>

# Thank you



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Slide 6: avatar icons, source: <https://iconscout.com/icon-pack/free-people-flat-design> (CC BY-SA 4.0, Author Coquet Adrien)