
University of
Southampton

University of Southampton

Technological Challenges of the DGA

and GDPR, Data Act, AI Act…

George Konstantinidis

New Approaches towards Compliance for AI/Data Operations
SINTEF, Oslo
Jan 2025

Data Governance Act

Effective on June 22, Applicable on Sep 23

Facilitate the sharing and reuse of data across the EU by

- creating a framework for data intermediaries,
- establishing safeguards for data sharing,
- and promoting the use of public-sector data for innovation.

Position: lack of incentives to participate in data sharing is the absence of ‘trust’.

This talk is not about:
Data Intermediaries vs Data Marketplaces

- Some data service providers want to sell data on their own
  - DGA advocates neutrality of DISPs, requires the legal separation of data intermediation and processing
- DISPs selling models instead of raw data?
  - processing data to train such models?
- DISPs offering data visiting rooms?
- DISPs do encrypted processing?
- Pricing?
- Auctions?

This talk is about:
Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

Citizen, Startup, SME

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

Citizen, Startup, SME

Public Body

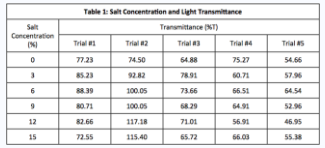


Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Data

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

Citizen, Startup, SME

Public Body

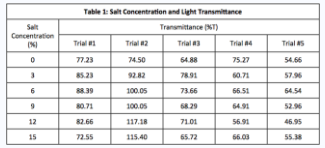


Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Data

IP holder, Data subject

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

Citizen, Startup, SME

Request: Purpose, Action

Public Body

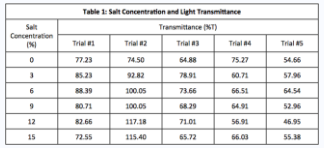


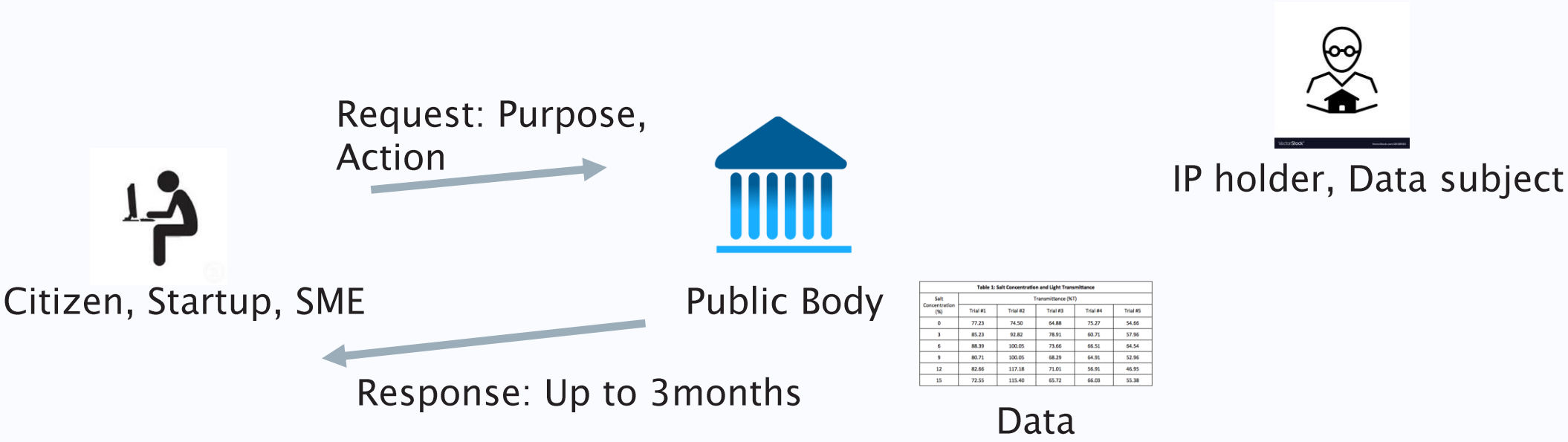
Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Data

IP holder, Data subject

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR



Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

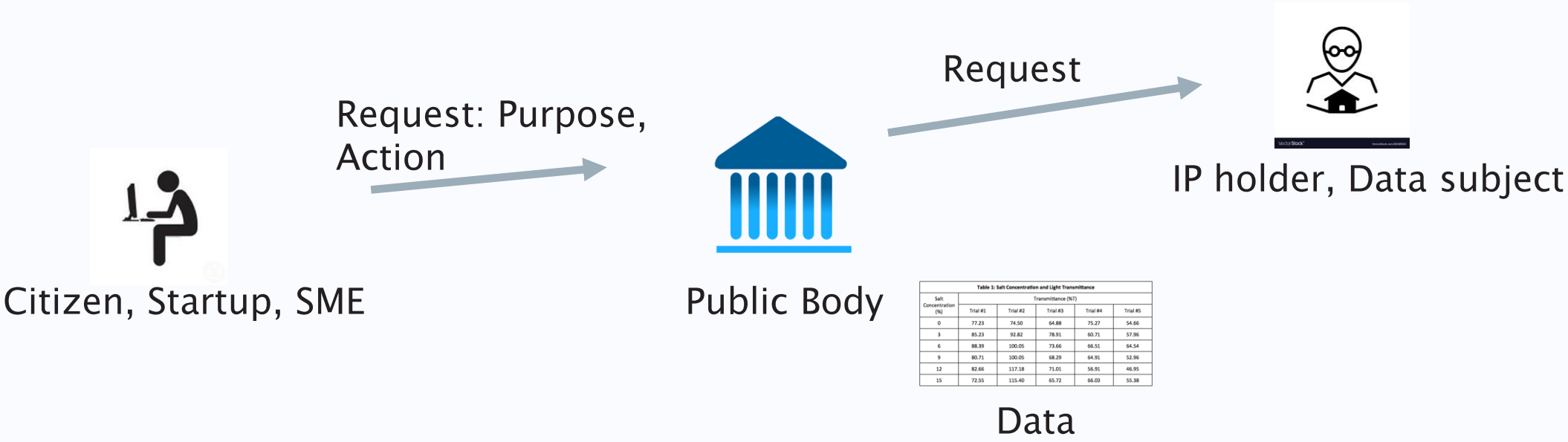


Table 1: Salt Concentration and Light Transmittance

Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Re-use of protected data held by public sector bodies

- Commercial data, statistical data, data protected by IP or protected personal data outside GDPR

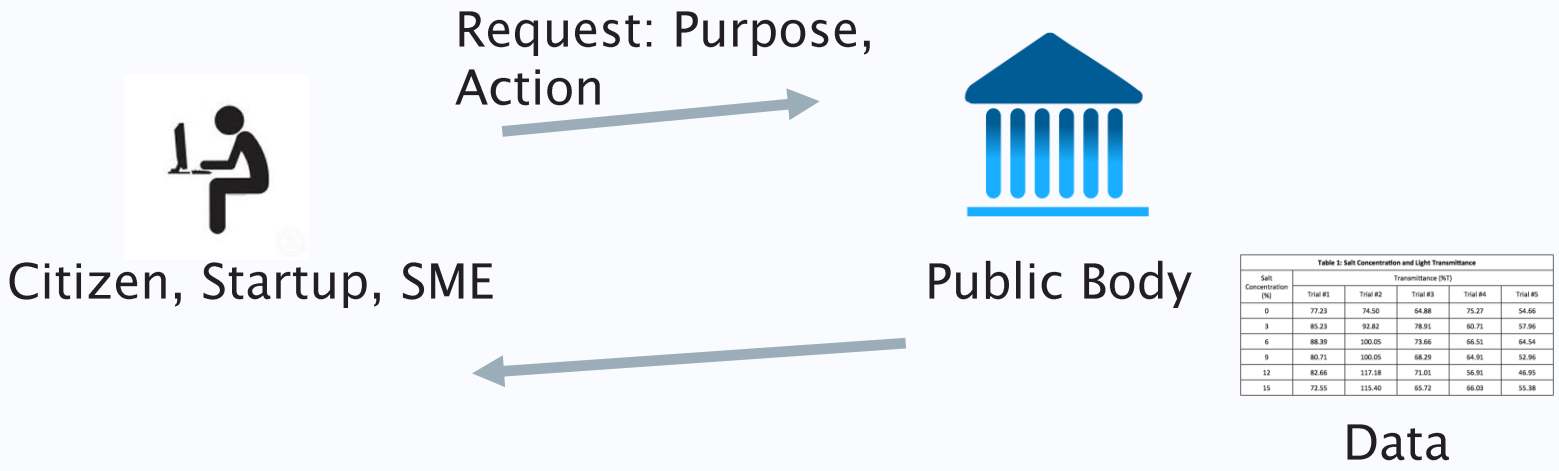


Table 1: Salt Concentration and Light Transmittance

Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

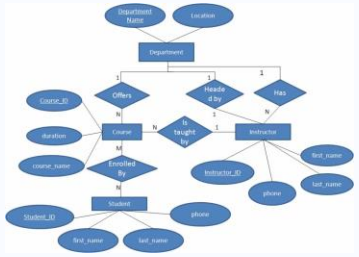
Granting Data Reuse

- Purposes of reuse
  - Agree on the same language
  - Non discriminatory, transparent, proportional
- Infrastructure to modify, anonymise, aggregate or any other method of disclosure control
  - Commercially confidential information
  - Trade secrets
  - Intellectual property rights
- Offer a secure processing environment
- Control Access
- Verify processing
- Monitor

Language of data use and reuse, e.g., “Purpose”

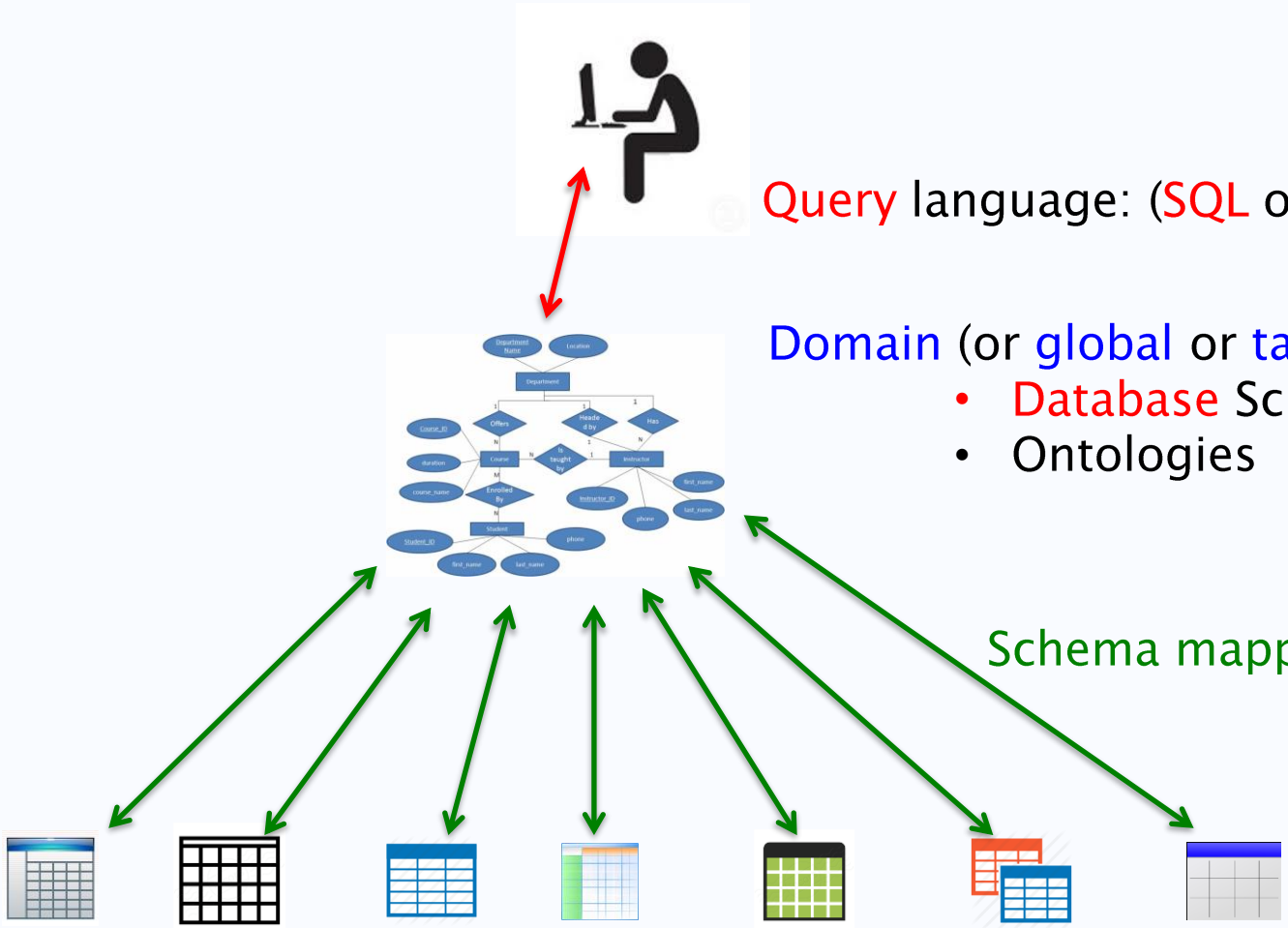
Purposes, Actions: Knowledge Graphs to the Rescue!

- High-level: advertisement, product research/enhancement, service improvement, security, diagnosis, etc.
- Medium level: Face recognition, age recognition, …, ?



- Low level: computational statistics, monte carlo simulations, linear regression, feature extraction, neural net training,…
- Domain-specific: biomedical techniques, healthcare choices,…

Logical Information Integration



Query language: (SQL or SPARQL)

Domain (or global or target) language:

- Database Schema
- Ontologies

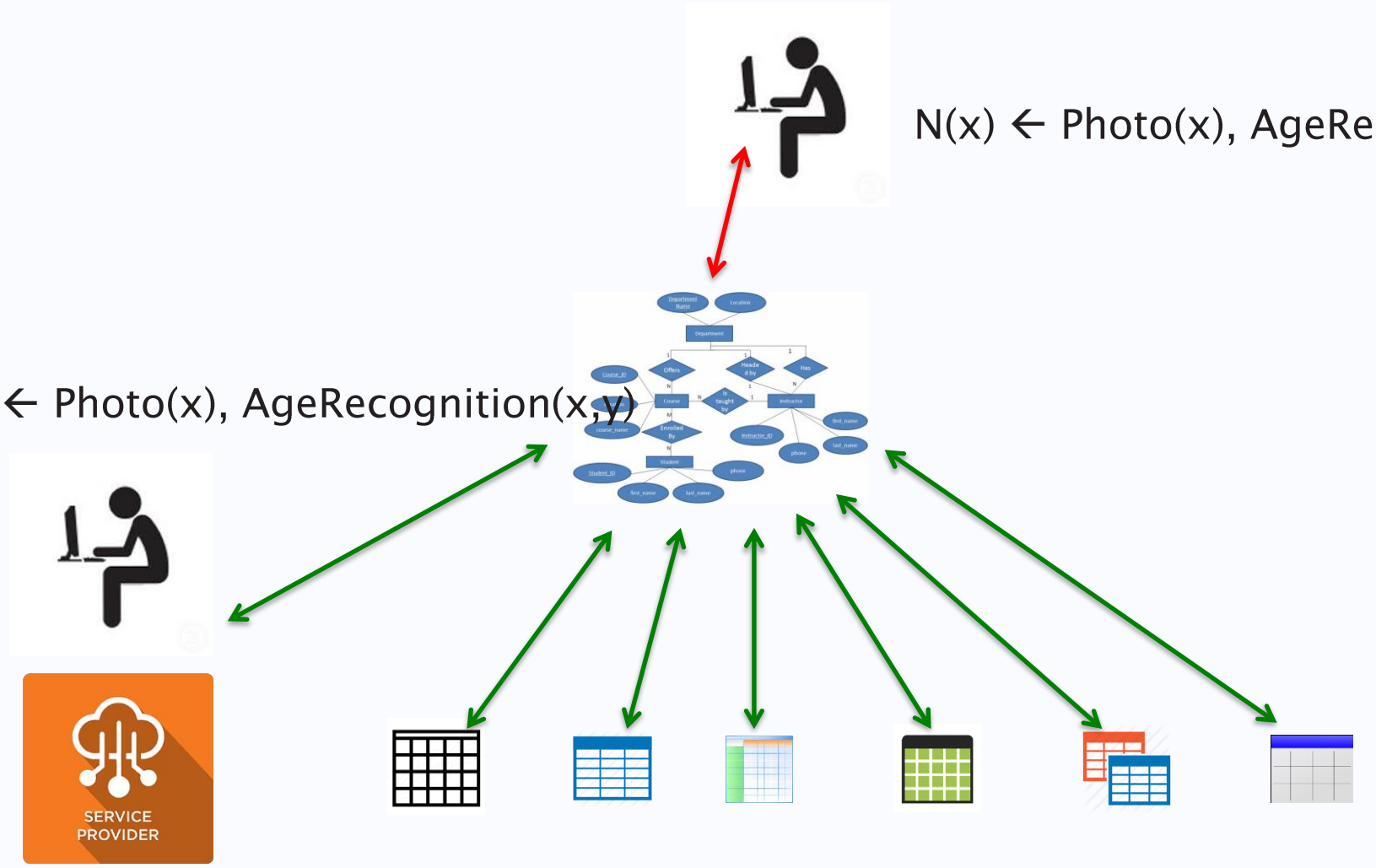
Schema mapping language:

Data Sources language

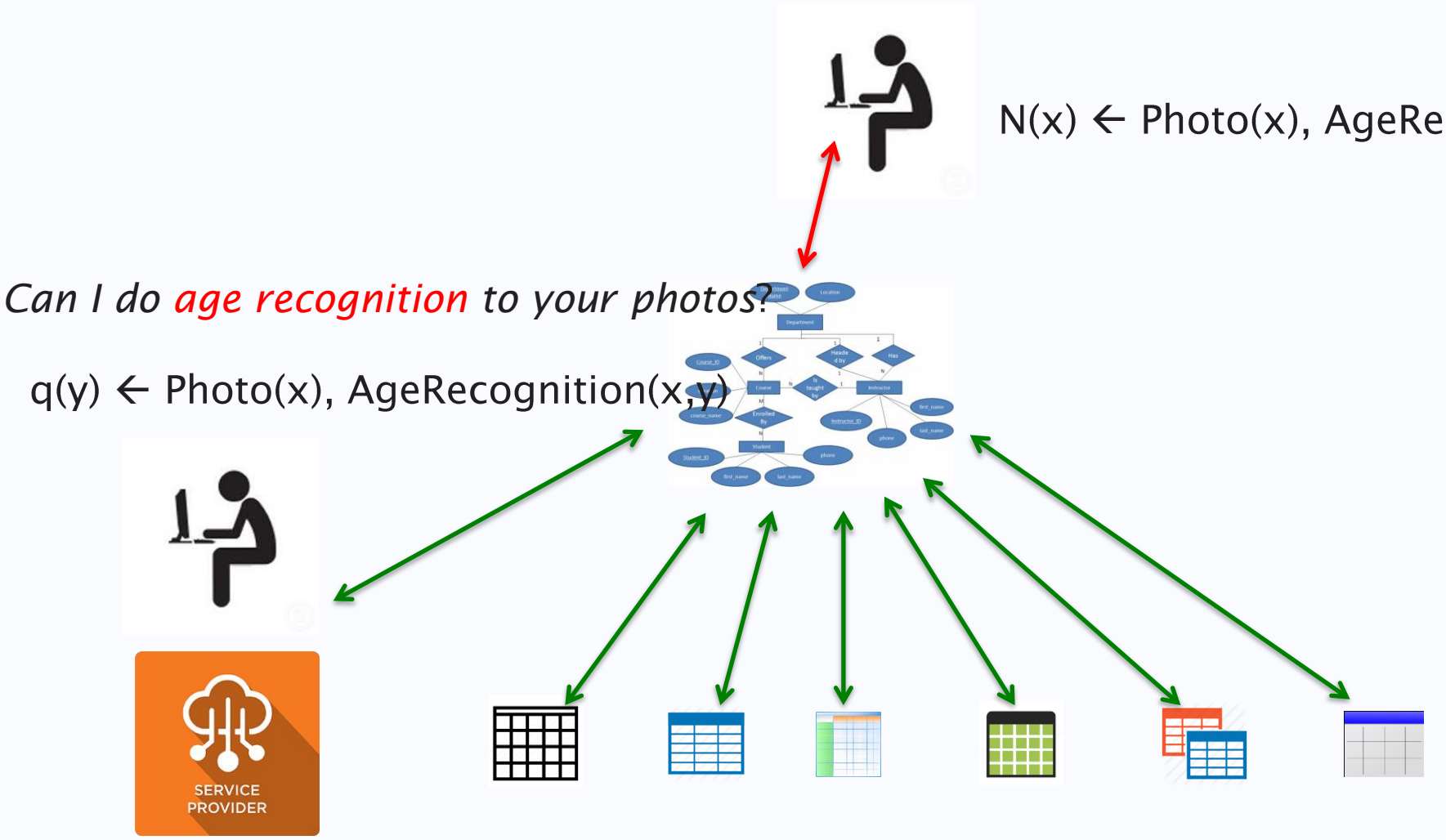
Integration of purposes and consent

N(x) ← Photo(x), AgeRecognition(x,y), y>50

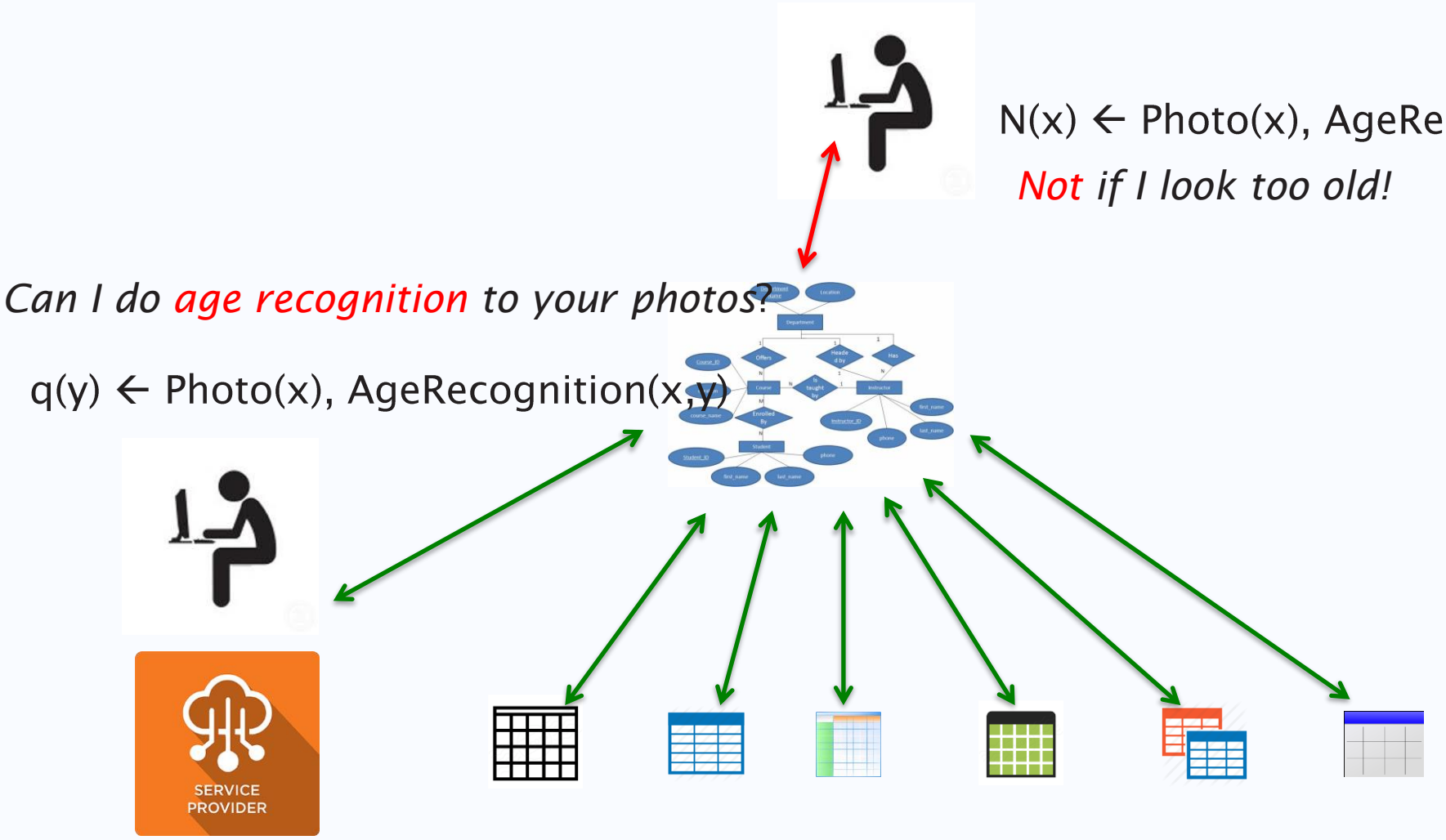
q(y) ← Photo(x), AgeRecognition(x,y)



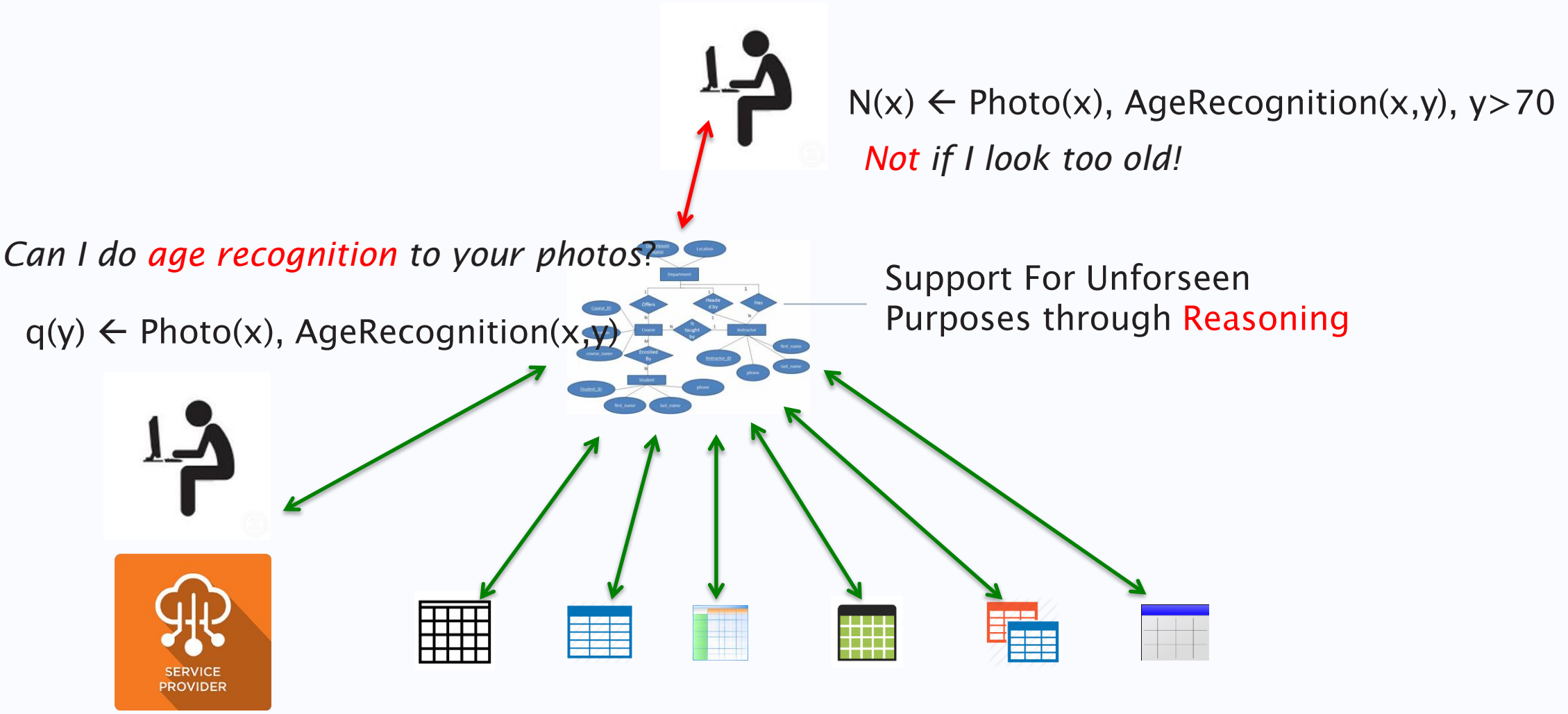
Integration of purposes and consent



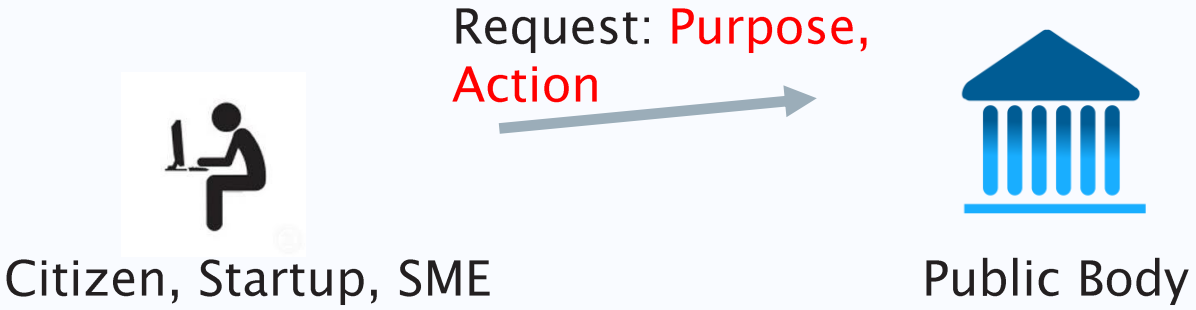
Integration of purposes and consent



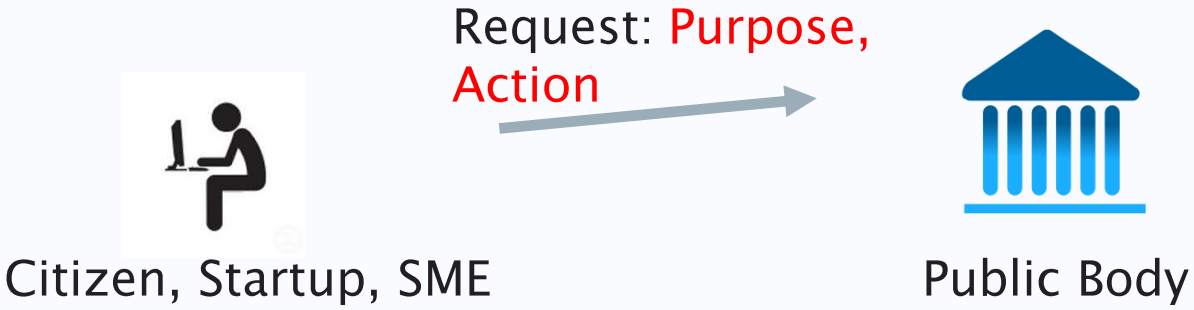
Integration of purposes and consent



Ok now I know what you say, but do you mean it?



Ok now I know what you say, but do you mean it? And who are you anyway?

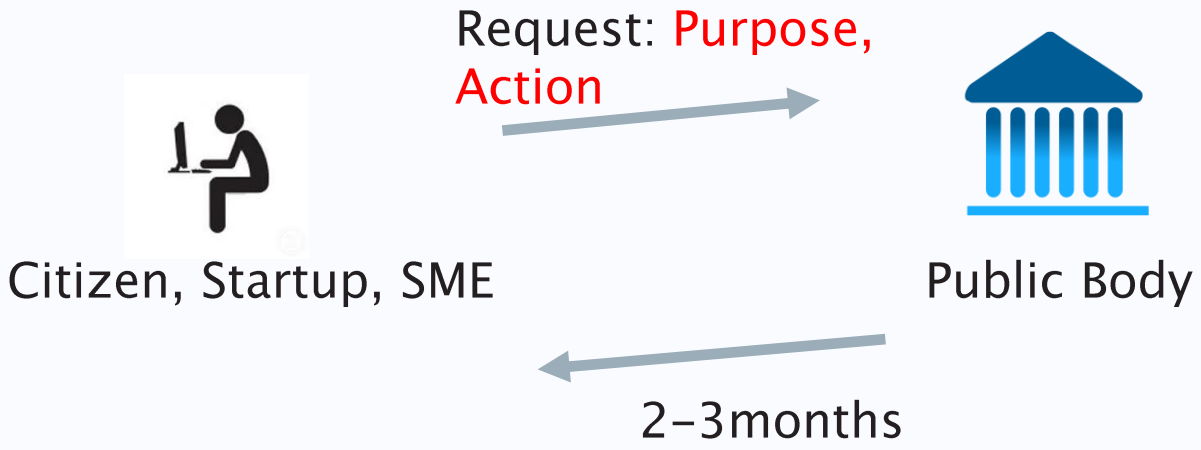


Trust

- Centralized Trust: Certification
  - Have organizations certify entities for purposes of use
  - You can do that a priori

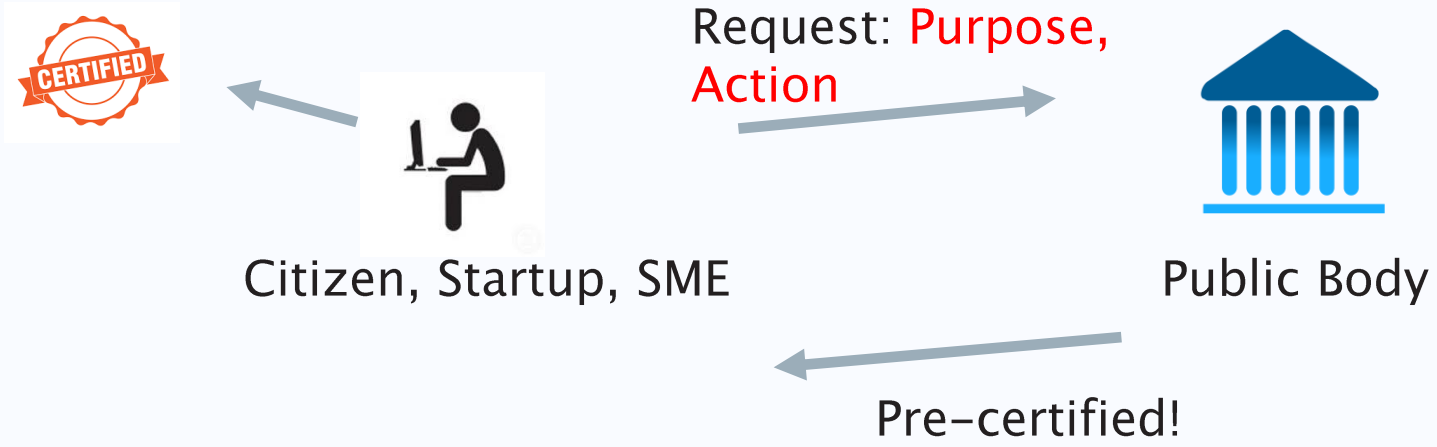
Trust

- Centralized Trust: Certification
  - Have organizations certify entities for purposes of use
  - You can do that a priori



Trust

- Centralized Trust: Certification
  - Have organizations certify entities for purposes of use
  - You can do that a priori

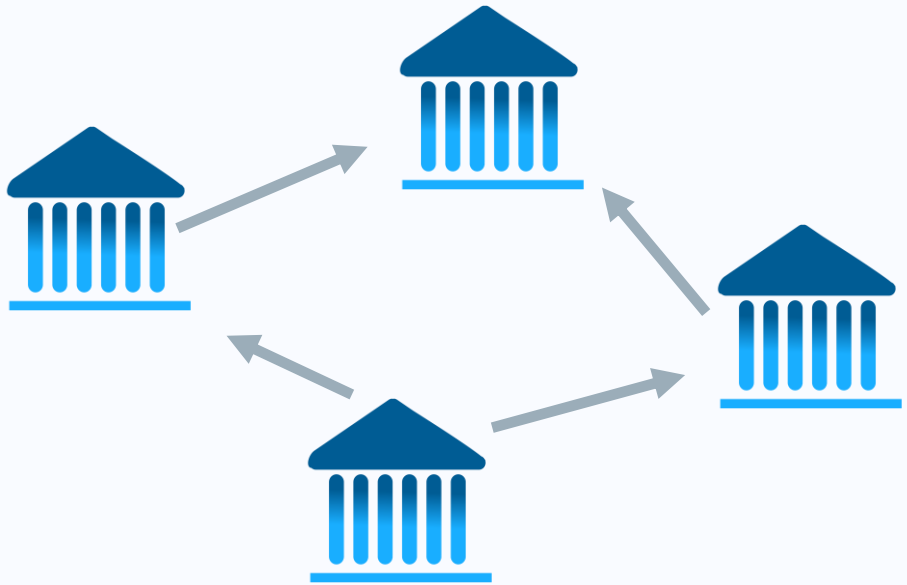


Trust

- Centralized Trust: Certification
  - Have organizations certify entities for purposes of use
  - You can do that a priori
  - Identity management
  - Verifiable credentials

Trust

- Centralized Trust: Certification
  - Have organizations certify entities for purposes of use
  - You can do that a priori
  - Identity management
  - Verifiable credentials
- Decentralized Trust



Secure Processing Environments

- Virtual or Physical
- Fine-grained policies
- Change Management
- Cost

Secure Processing Environments

- Virtual

Request: Purpose, Action

Public Body

Send me your algorithm

Data Room

Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Secure Processing Environments

- Virtual

Request: Purpose, Action

Public Body

Send me your algorithm

How?

Data Room

Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Secure Processing Environments

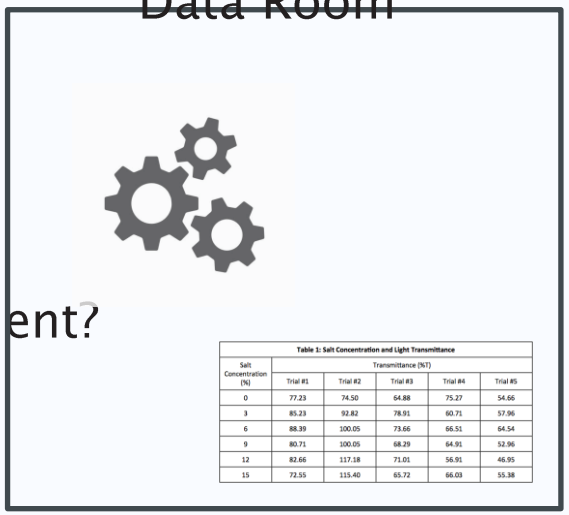
- Virtual

Request: Purpose, Action

Public Body

Data Room

- System configuration
- Synthetic data
- Remote algorithm development?



Secure Processing Environments

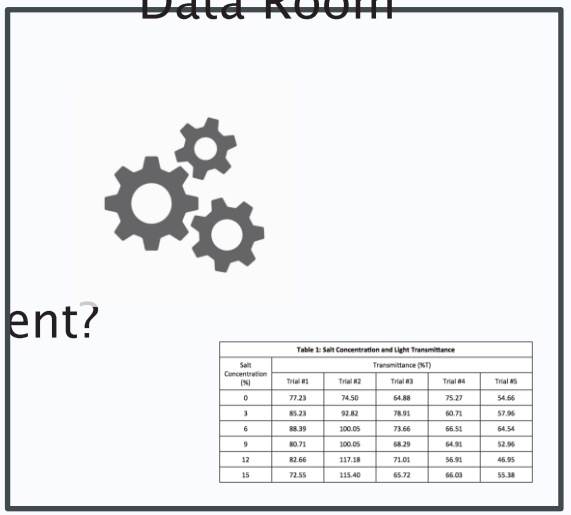
- Virtual

Request: Purpose, Action

Public Body

Data Room

- System configuration
- Synthetic data
- Remote algorithm development?
- Virtual sandbox



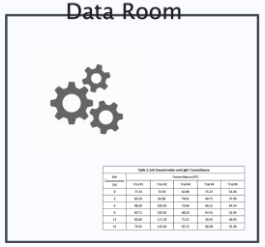
Secure Processing Environments

- Virtual

Request: Purpose, Action

Public Body

Data Room



Data Room

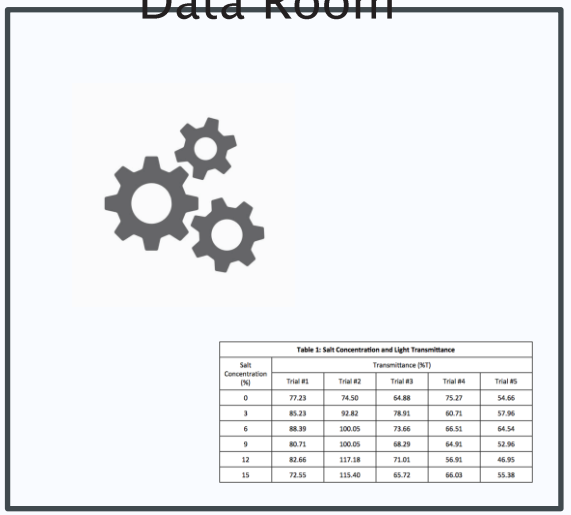
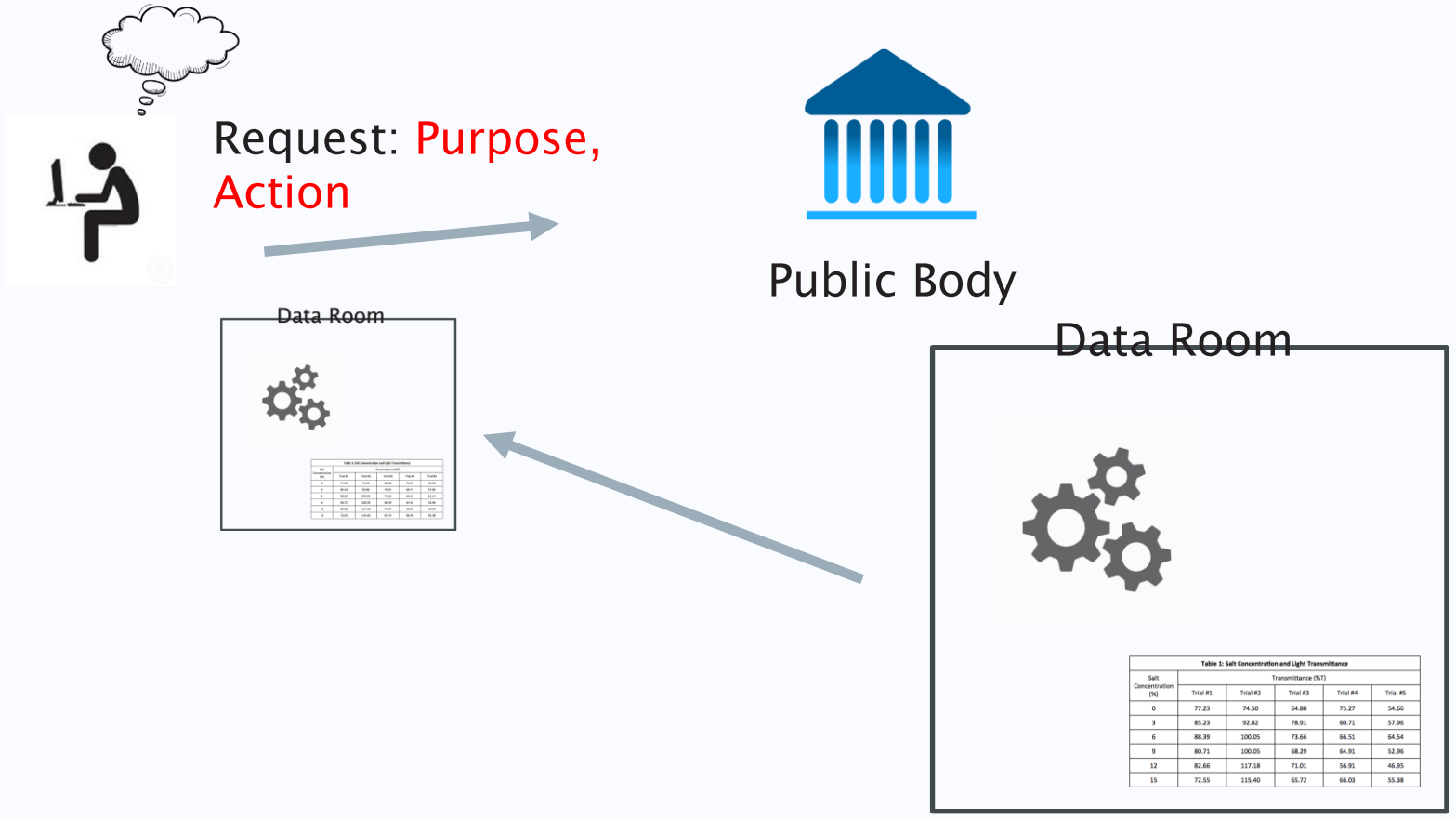


Table 1: Salt Concentration and Light Transmittance					
Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

Secure Processing Environments

- Virtual



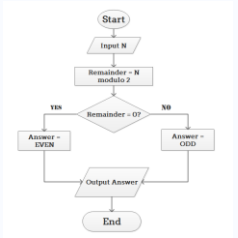
Secure Processing Environments

- Virtual

Request: Purpose, Action

Public Body

Algorithm



Data Room

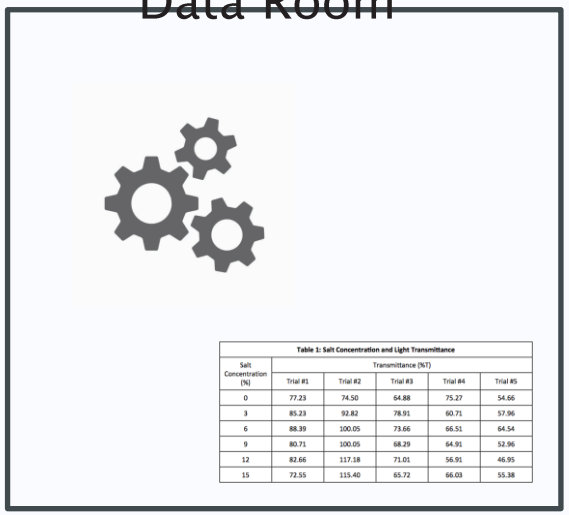


Table 1: Salt Concentration and Light Transmittance

Salt Concentration (%)	Transmittance (%T)				
	Trial #1	Trial #2	Trial #3	Trial #4	Trial #5
0	77.23	74.50	64.88	75.27	54.66
3	85.23	92.82	78.91	60.71	57.96
6	88.39	100.05	73.66	66.51	64.54
9	80.71	100.05	68.29	64.91	52.96
12	82.66	117.18	71.01	56.91	46.95
15	72.55	115.40	65.72	66.03	55.38

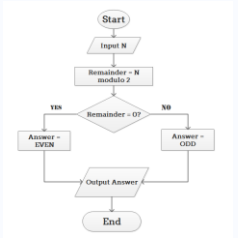
Secure Processing Environments

- Virtual

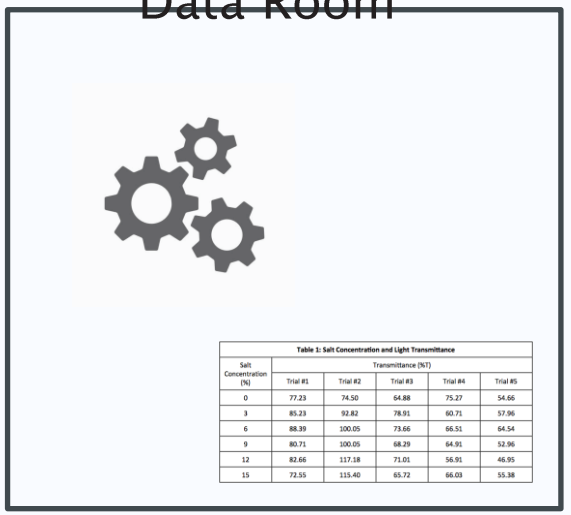
Request: Purpose, Action

Public Body

Algorithm



Data Room



Monitoring
Verify
Audit

a trusted party?

How about protecting against an ~~adversary~~

Data to be released

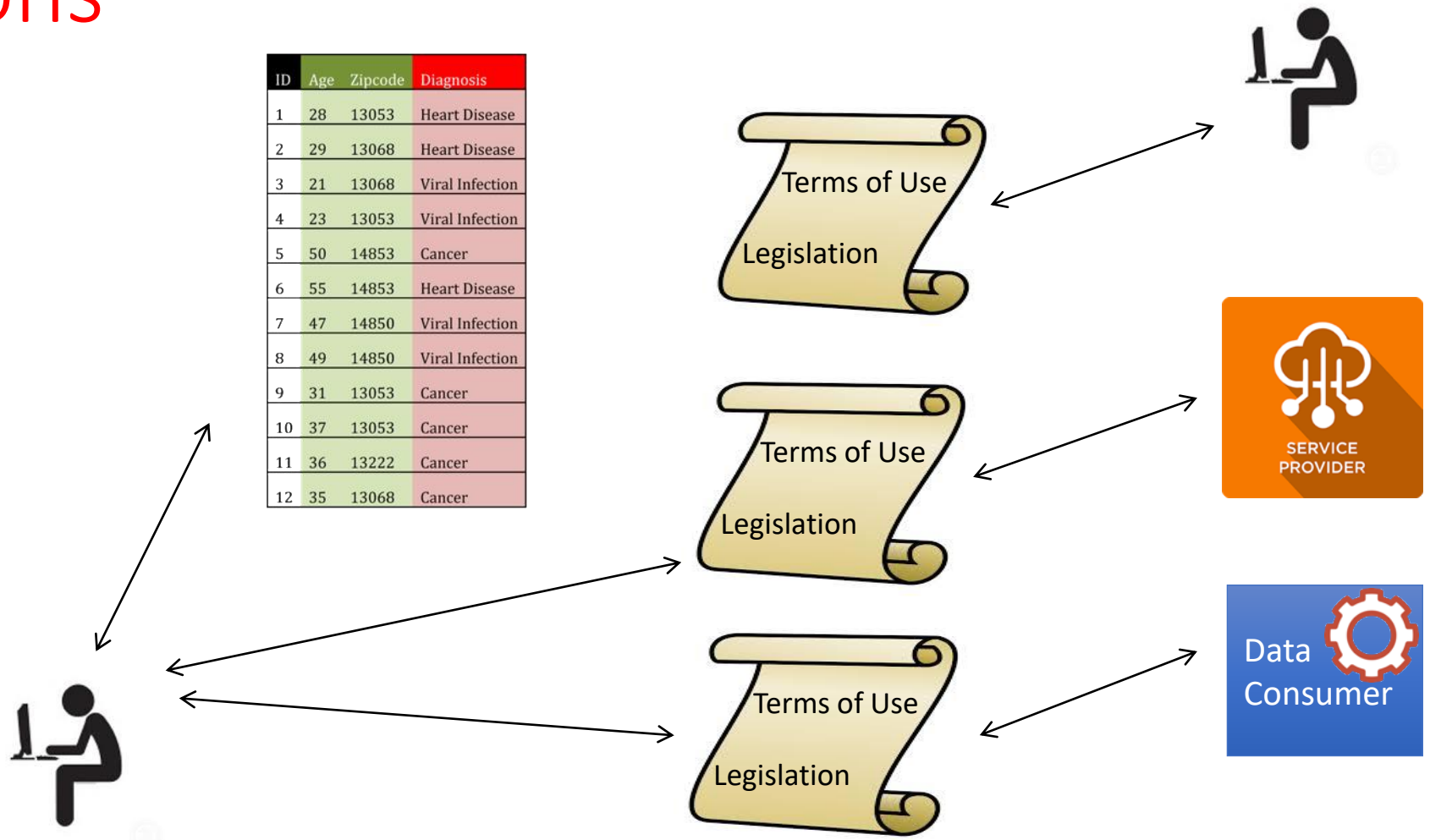
- Data access
- Encryption
- Anonymisation
- Differential Privacy
- …

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

Data Owner

Data Agreements Currently: Terms and Conditions

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer



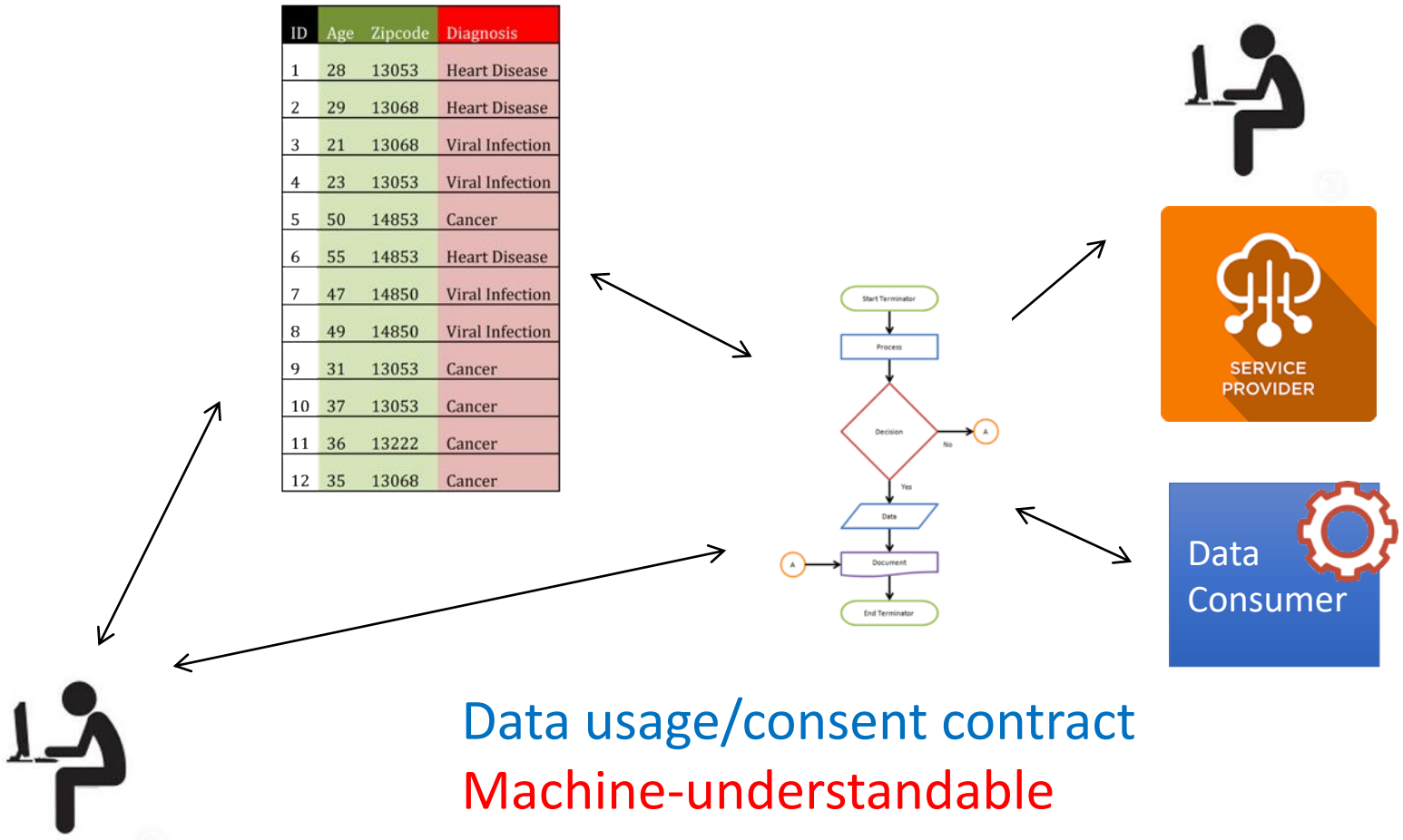
Needs in-house engineers to do an ad-hoc implementation of the terms into code

Data Agreements Currently: Terms and Conditions

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

Can we do better?

Collaborative Privacy over released data



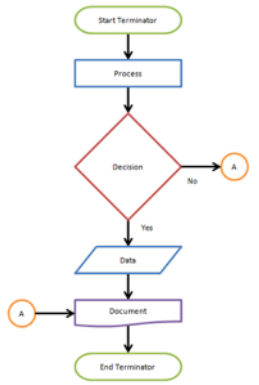
Data usage/consent contract

Machine-understandable

Collaborative Privacy over released data

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

Fine-grained



Data usage/consent contract

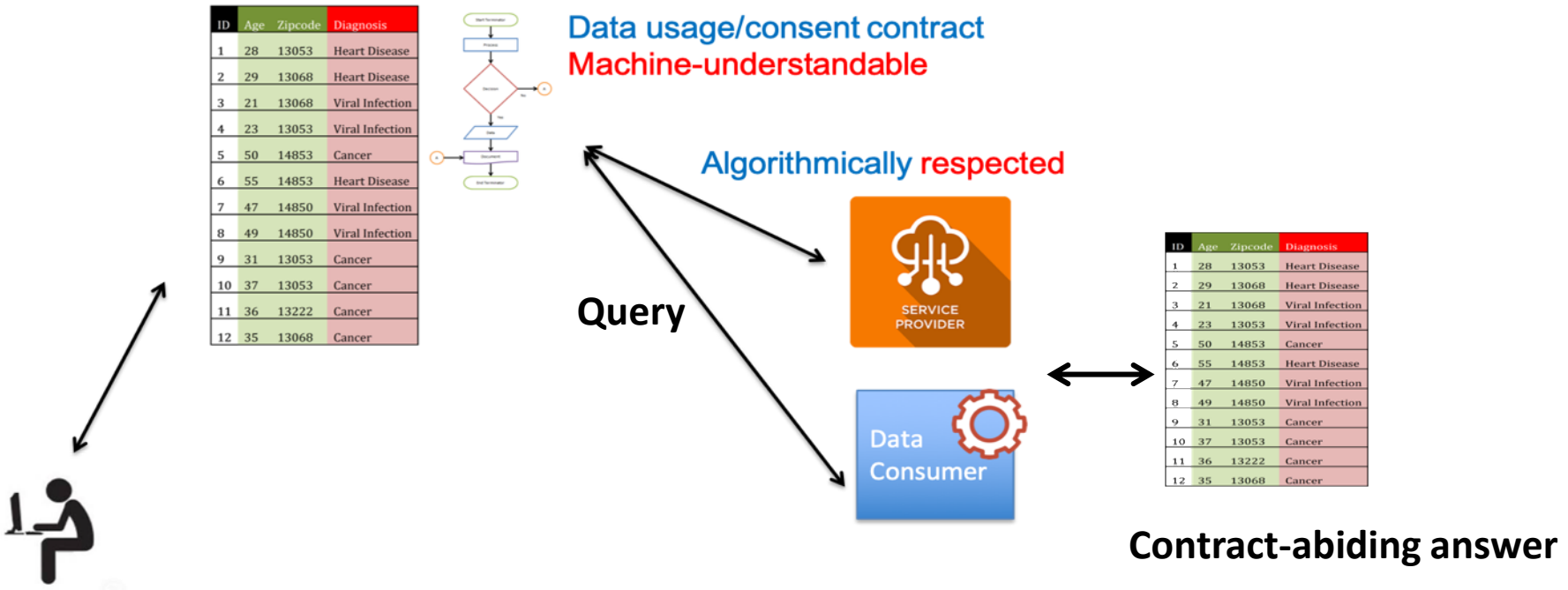
Machine-understandable

Bottom-up

Algorithmically respected

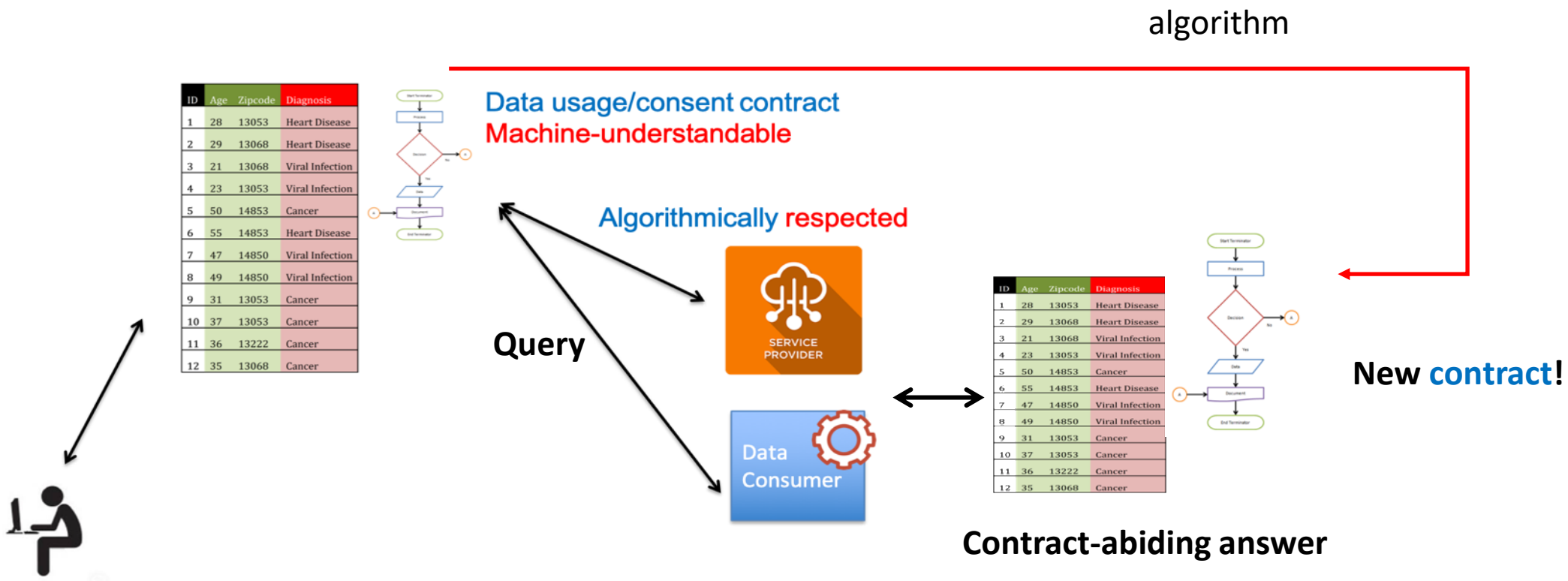
Data Sharing Agreements

- More contracts automatically generated for the processed data



Data Sharing Agreements

- More contracts automatically generated for the processed data



Data Altruism

- Not enough time!

University of Southampton

YOUR QUESTIONS

g.konstantinidis@soton.ac.uk