Universiteit van Amsterdam
Instituut voor Informatica

Deep Tech Day -- Cybersecurity

Compliant Data Usage in Cyberinfrastructure

Dr. L.T. (Thomas) van Binsbergen
Associate Professor
l.t.vanbinsbergen@uva.nl

Security research at Informatics Institute

- Systems Security (Complex Cyber Infrastructure group)
  - Secure Data Exchange
    - Enforcement of regulations and agreements
    - Dynamic workflow adaptations for on-going compliance
  - Privacy Enhancing Technologies (PETs)
    - Secure Neural Network Inference, Energy efficiency
    - Federated Machine Learning, Policy-driven
- Hardware Security (Parallel-Computing Systems group)
  - Side-channel and fault attacks,
  - secure SW/HW implementations
- Quantum and post-quantum Cryptography (Theoretical Computer Science group)
  - Secure Multi-Party Computation protocols (sMPC), Feasibility and Efficiency

Personal Introduction

- Dr. L.T. (Thomas) van Binsbergen
  Associate Professor, Informatics Institute (Lab42)
  Complex Cyber Infrastructure (CCI) group

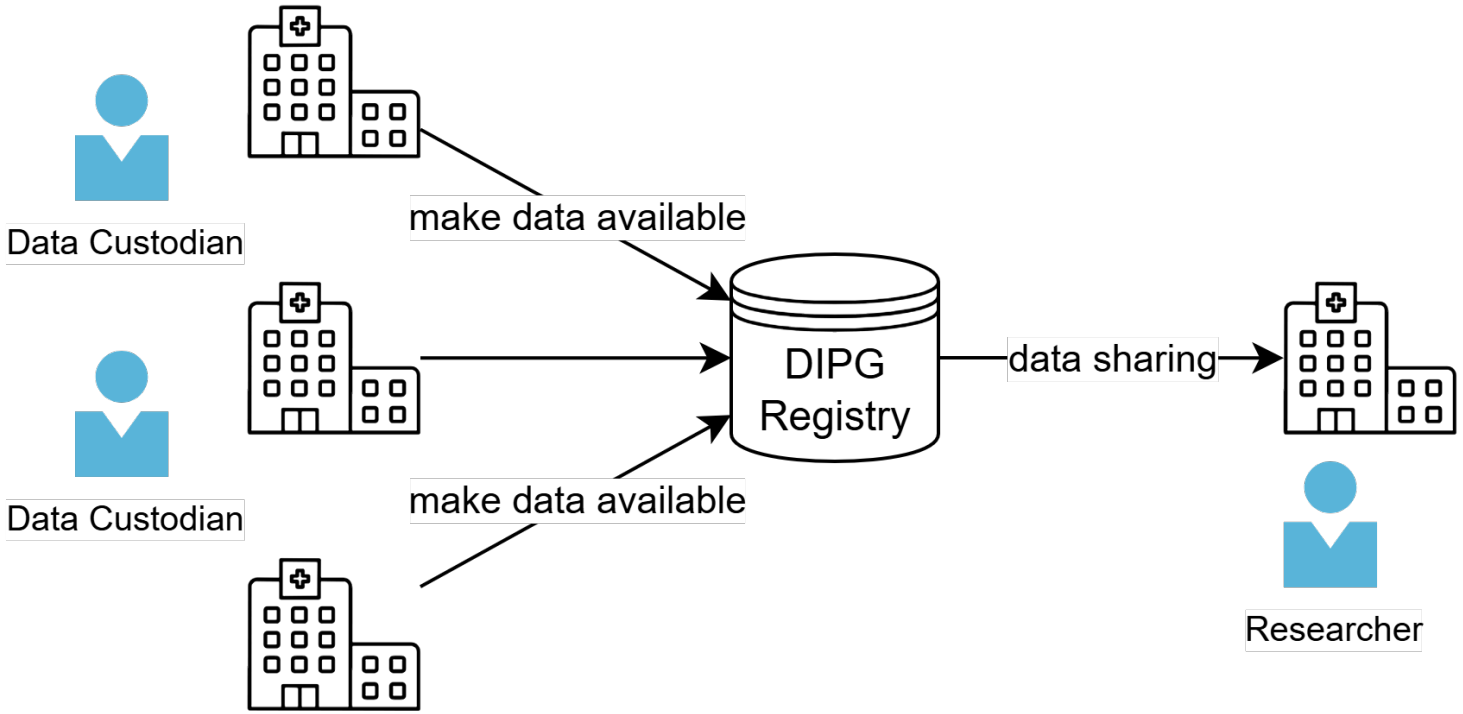
- Expertise: Software Engineering, Programming Languages, Domain-Specific Languages (DSLs), Compliance in Software

- Advocate of ‘Rules as Code’. Formalising legal rules as executable code and integrating this code in software systems

- Application domain: data exchange in cyberinfrastructure, leveraging security mechanisms for compliant data exchange

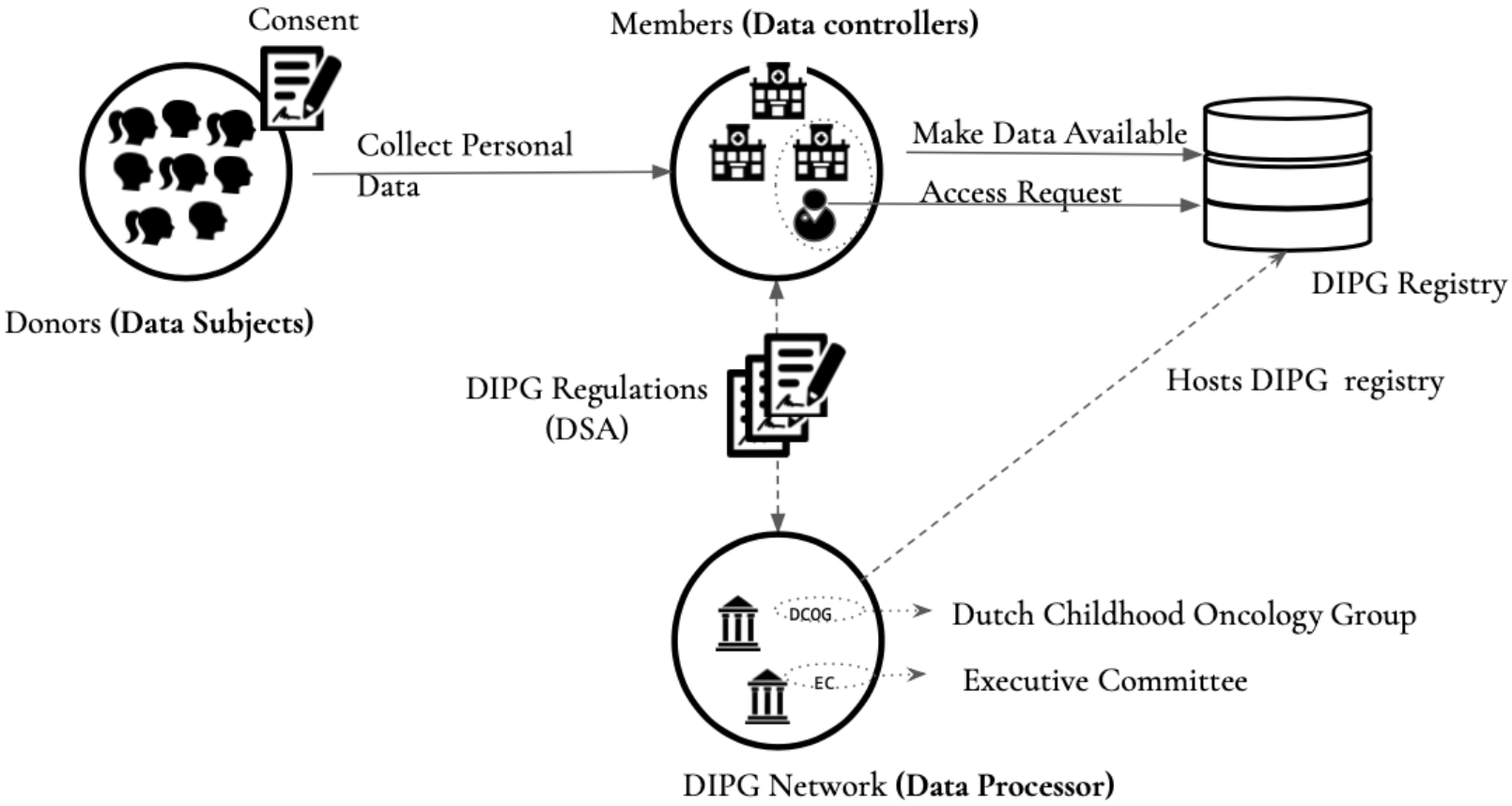
Multi-Organisational Data Sharing (examples)

- **Goal**: Establish neutral and domain-agnostic infrastructure for the exchange of data, enabling data-service providers, whilst enforcing agreements, laws, and regulations.
  - Fosters data- and AI-driven innovation.



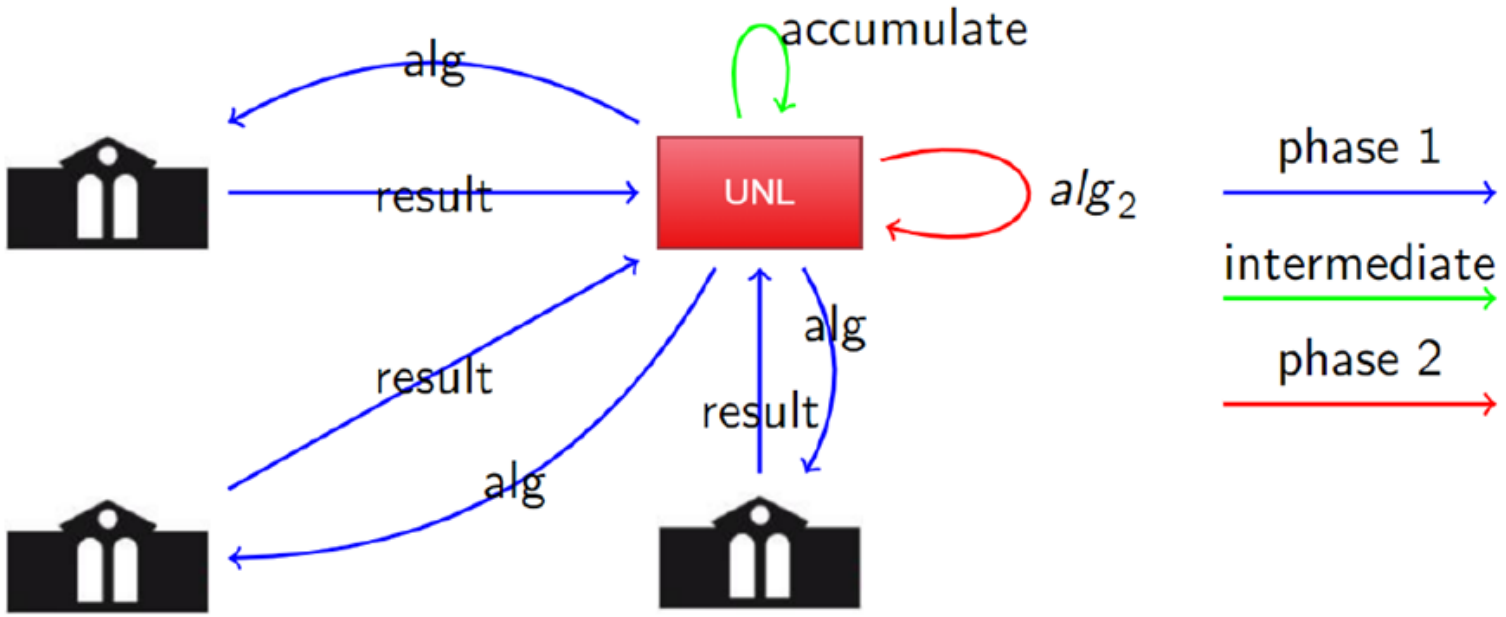
Medical data sharing via a central registry (example: DIPG network)

Multi-Organisational Data Sharing (examples)



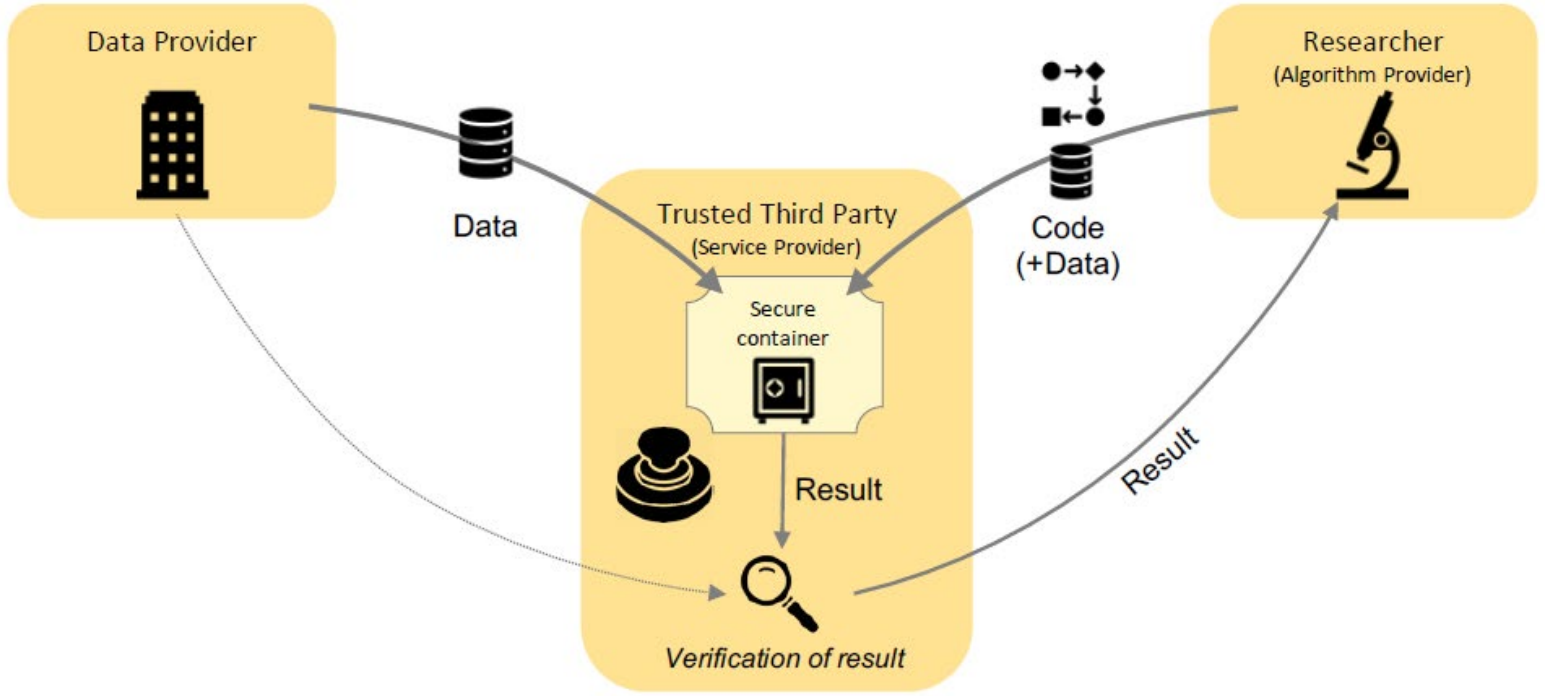
Data Governance architecture of previous DIPG example

Multi-Organisational Data Sharing (examples)



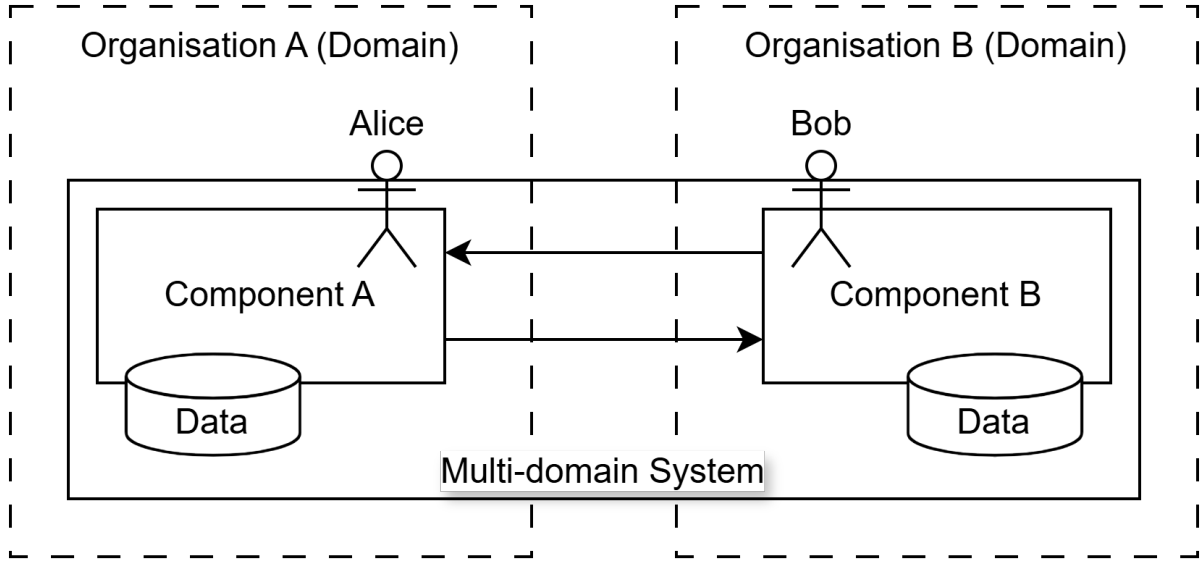
Multi-party Compute-to-Data processing (example: University personnel files)

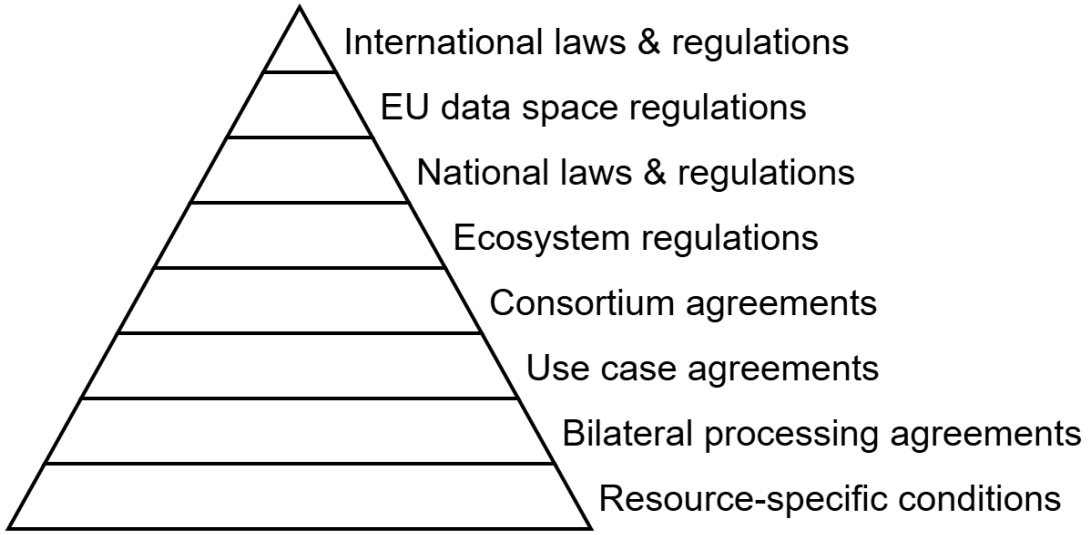
Multi-Organisational Data Sharing (examples)



Compute-to-Data processing via Trusted Third Party (example: SANE by SURF)

Control in Multi-Organisational Software





- How can Alice control whether to send data to Bob? (access control)
- How can Alice control how Bob can use the data A sends to B? (usage control, post-duties)
- How can Alice ensure that A receives from B what she expects? (duty enforcement)

  Proposed solution:

  Laws, contracts, usage conditions formalised and enforced as compliance specifications

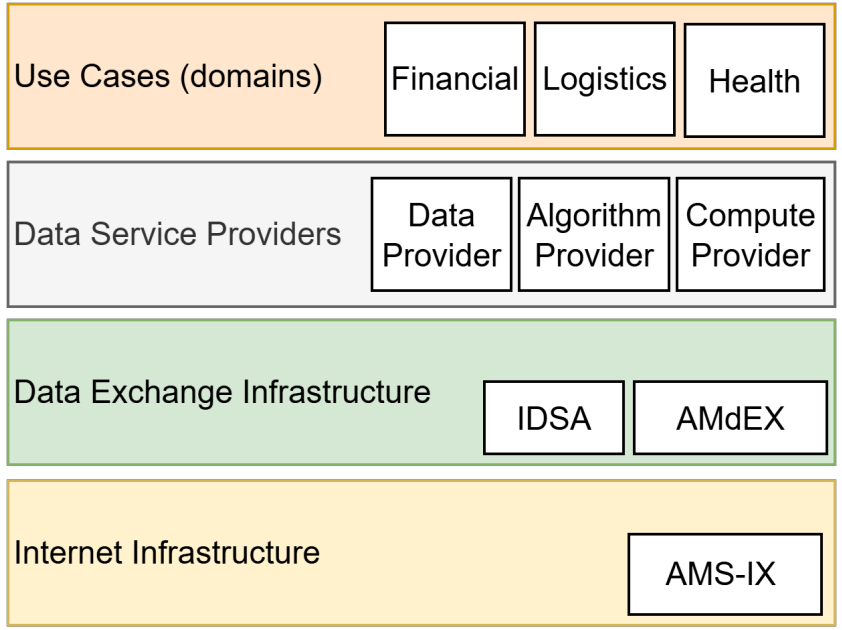
  → how to distribute and decentralise this kind of enforcement?

The Amsterdam Data Exchange (AMdEX)

Vision and initiatives

- **Goal**: Establish neutral and domain-agnostic infrastructure for the exchange of data, enabling data-service providers, whilst enforcing agreements, laws, and regulations.
  - Fosters data- and AI-driven innovation.
- **Method**: Use-case driven, bottom-up design of (technical) roles, processes and components with open-source implementations.

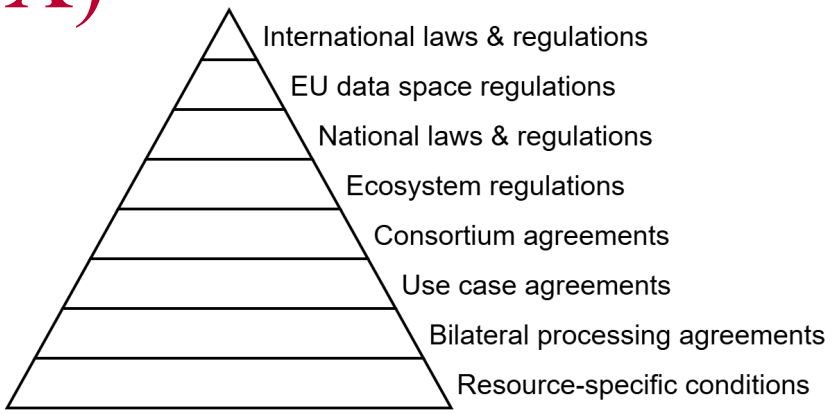
- **Initiatives**:
  - Fundamental research in several NWO projects (2018-2023)
  - Fieldlab with SMEs and societal partners (2020-2023)
  - Operationalisation in the DMI national ecosystem for (2024-)
- **Results**:
  - AMdEX Reference Architecture v1
  - 10+ PoCs and Demos (TRL3-6) across research projects
  - Components operational (TRL7+) within DMI ecosystem
  - Compatible with IDSA reference architecture

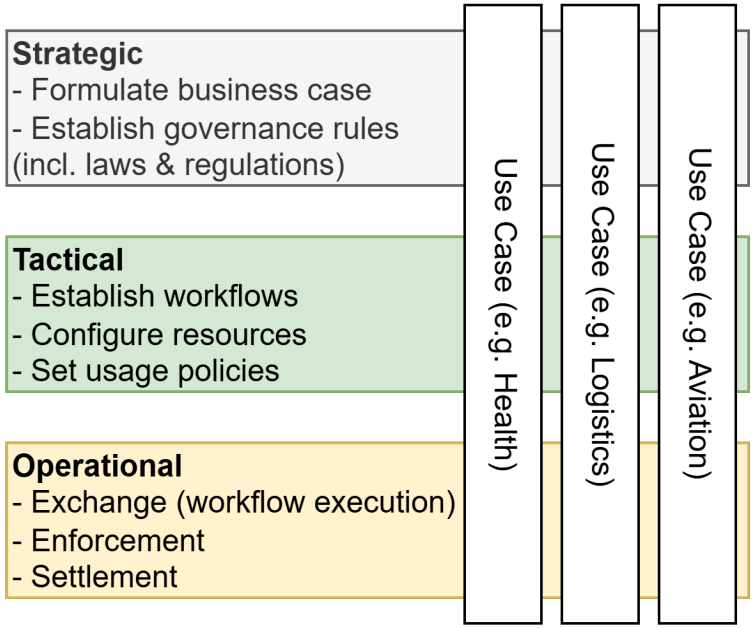


The Amsterdam Data Exchange (AMdEX)

High-level solution ingredients

- **Domain-specific language** for the formalisation of laws, regulations, agreements, resource-specific conditions, and their concretisation as technically *enforcable policies*.
- **Demand-driven** consortium formation, connecting infrastructure- and service-providers according to common solutions defined as *data exchange archetypes*.
- **Policy-driven orchestration** and **dynamic adaptation** scheduling tasks across participants to execute *workflows*.
- **Settlement and auditing** supported by monitoring and event logs.





Identification, Authentication, Authorization (IAA)

- **Identification**: the *claim* of who you are.
  For example, I can claim to be the person 'owning' l.t.vanbinsbergen@uva.nl
  The email address is the identity I present.

  Analogy: showing an ID card to security.

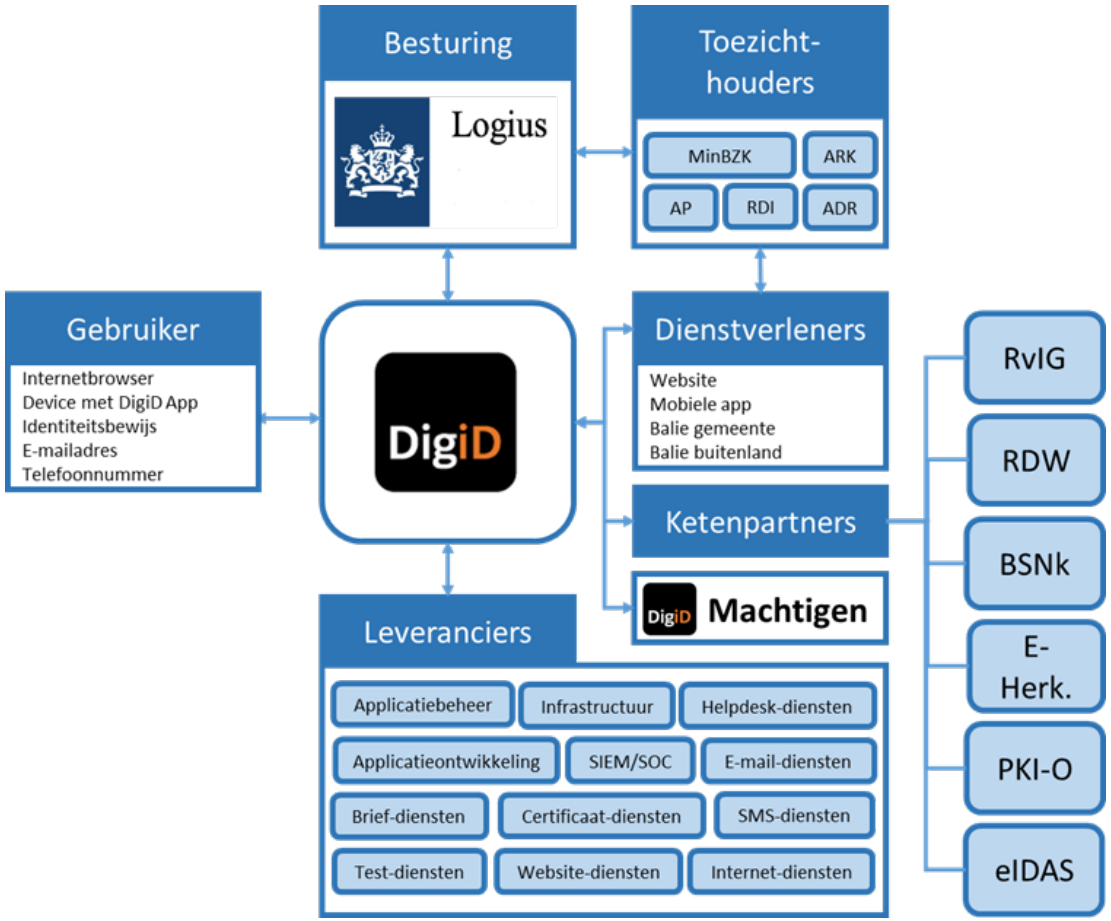
- **Authentication**: the *proof* that you are who you claim.
  For example, I can demonstrate ownership of my identity by entering a verification code sent by email.
  Methods like two-factor authentication (2FA) strengthen authentication (as the name suggests).

  Analogy: the security guard compares the picture on the card with your face.

- **Authorization**: the *set of things* you are allowed to do once your identity is established.
  For example, I can control certain project budgets (but crucially not all).
  Authorization is dynamic, granular, context-sensitive, case-specific, etc. etc.

  Analogy: security grants access to floors 1-9, but not 10-15.

DigiD



Betrouwbaarheidsniveaus uit de Europese eIDAS-verordening

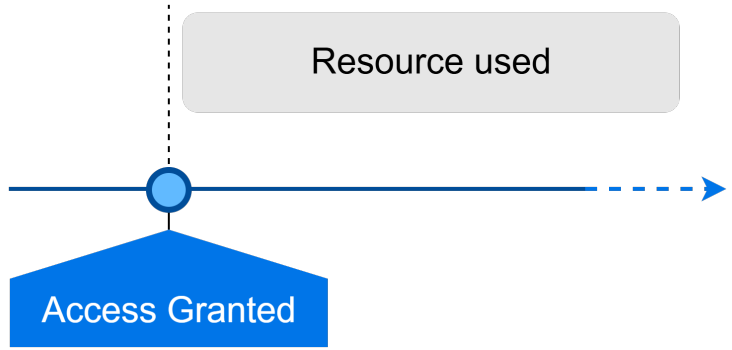
DigiD-niveau	eIDAS-niveau	Omschrijving	Status
Basis	Laag	1-factor-authenticatie met gebruikersnaam en wachtwoord	Wordt uitgefaseerd
Midden	Laag	2-factor-authenticatie met sms of app	Blijft beschikbaar
Substantieel	Substantieel	Eenmalige ID-check in de app	Wordt gestimuleerd
Hoog	Hoog	ID-check bij elke inlog via de app	Wordt doorontwikkeld; nog niet toegepast door dienstverleners

DigiD wordt beheerd door de Nederlandse overheidsorganisatie Logius, onderdeel van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Het platform waar DigiD op draait wordt geleverd door Solvinity en *draait in een overheidsdatacentrum*. Dit maakt *Solvinity een leverancier* van DigiD, niet de eigenaar of ontwikkelaar van de software. – https://www.digid.nl/solvinity

Wat zijn de diensten van Solvinity?

- Garanties op hoge mate van beschikbaarheid
- Hosting en diensten voor beheer van infrastructuur
- Diensten voor het monitoren van gebruik (o.a., voor security doeleinden)

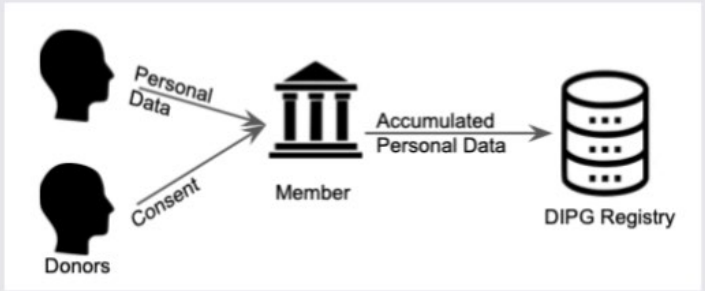
Authorization: Access Control



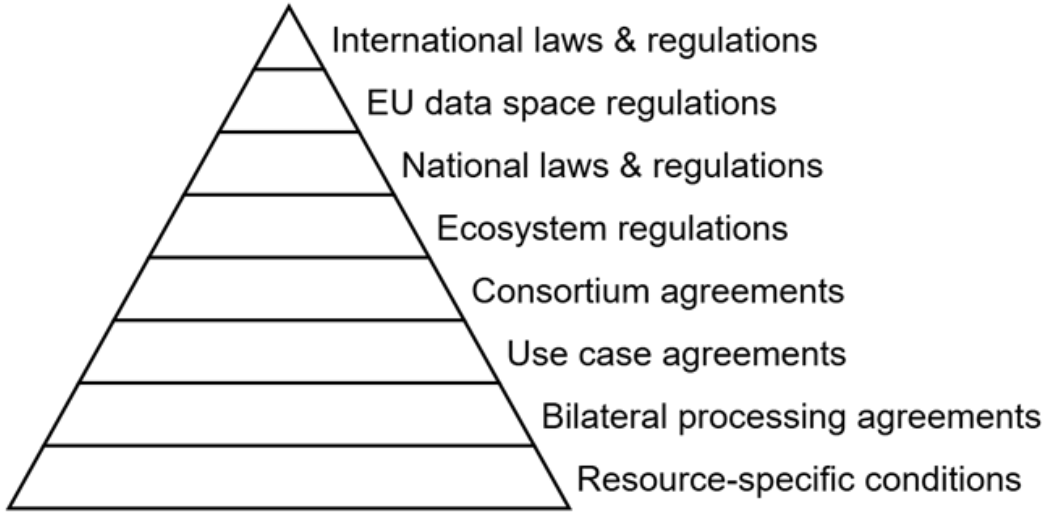
May actor X access asset Y to perform action read/write?

Question 1

What conditions need to be fulfilled before making data available?

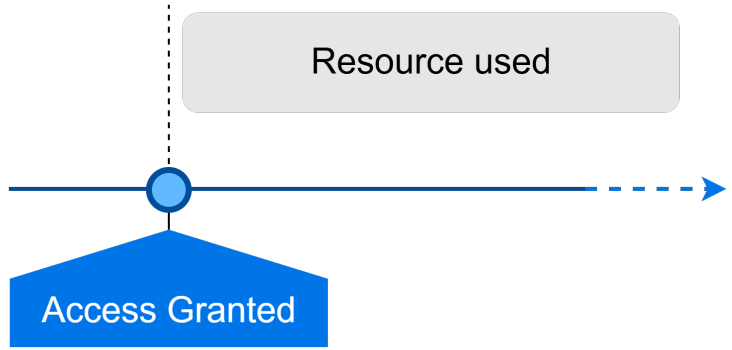



```
?Enabled(write(<X>,<Y>))
```



Result: *codified* GDPR articles, DIPG articles and consent statements are *translated* and enforced as *system-level access control policies*

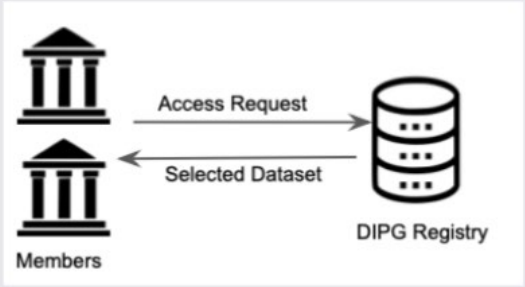
Authorization: Access Control



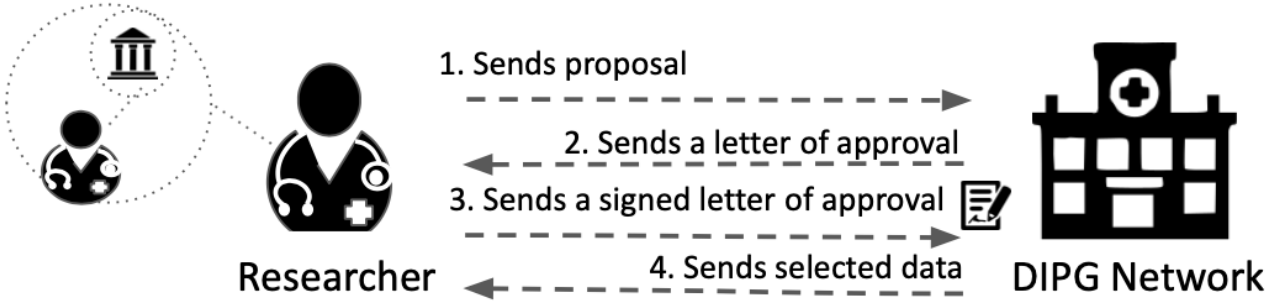
May actor X access asset Y to perform action read/write?

Question 2

What conditions need to be fulfilled when accessing data from the registry?



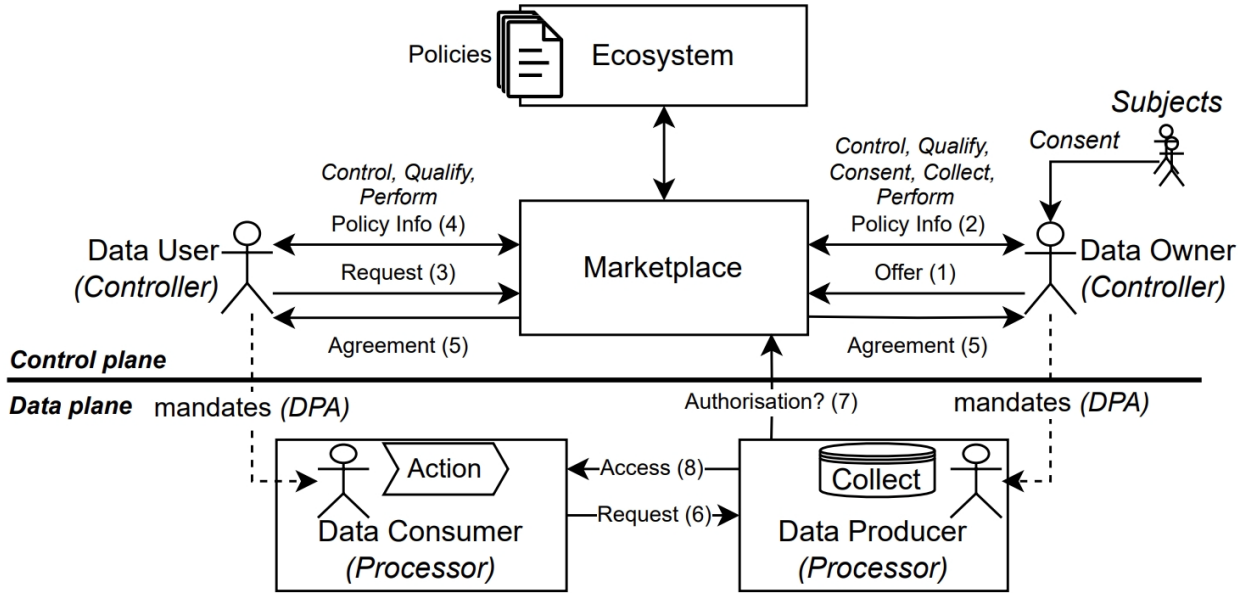

```
?Enabled(read(<X>,<Y>))
```



Result: *codified* GDPR articles, DIPG articles and consent statements are *translated* and enforced as *system-level access control policies*

Purpose-based Access Control

- Purpose ("doelbinding") is a central concept in GDPR and other regulations related to data processing
- Authorisations are given only when Data User processes data for acceptable purposes



DMI·ECOSYSTEEM

The DMI ecosystem is a collaboration of companies, knowledge institutes, G40 and G4 municipalities, provinces and the Ministries of IenW en BZK made possible in part by the National Growth Fund.

Technical and data governance architecture for personal data within the DMI ecosystem. A Data Owner and User come to an agreement on processing purposes based on which a Data Consumer and Data Producer can be authorised for peer-to-peer data sharing

From Access Control to Usage Control

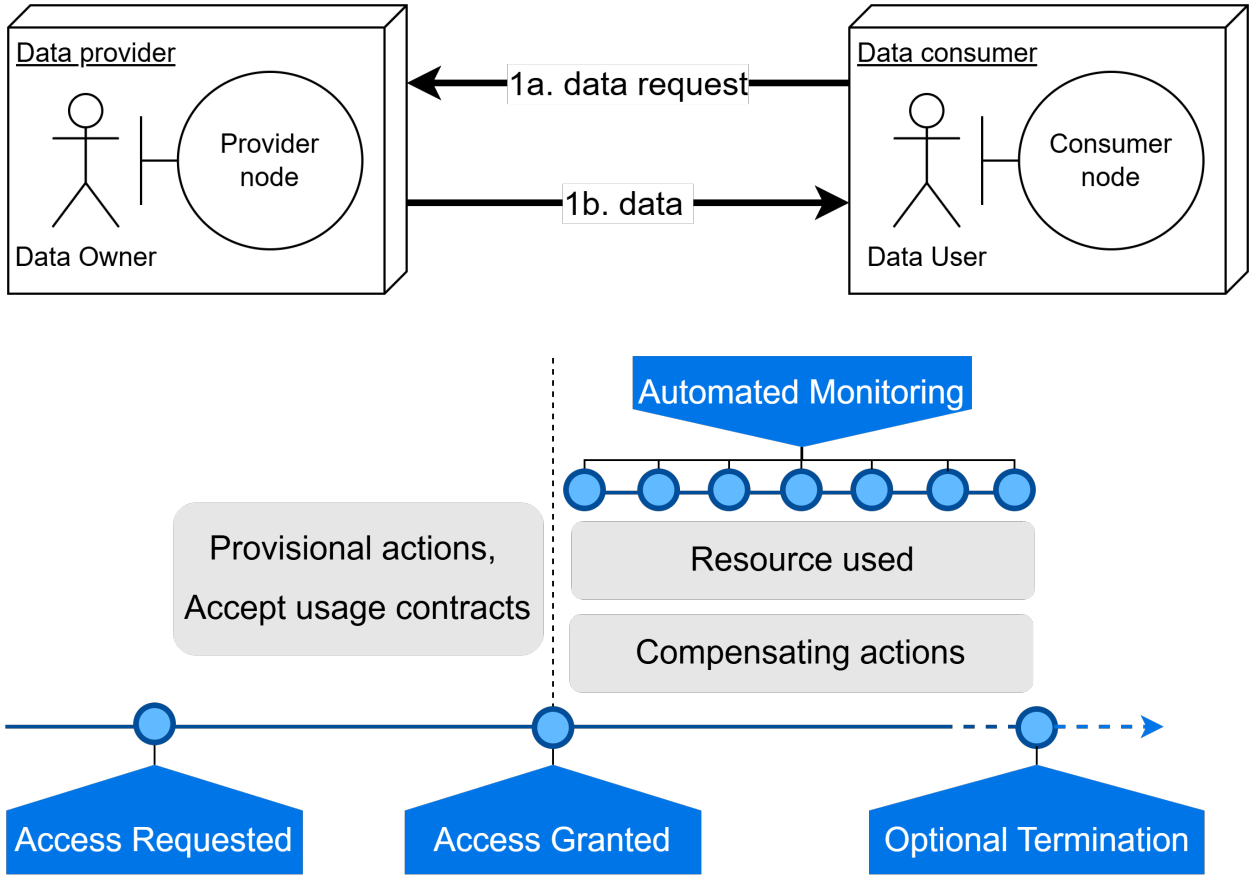


Image adjusted from "Distributed Usage Control – Pretschner et al." https://doi.org/10.1145/1151030.1151053

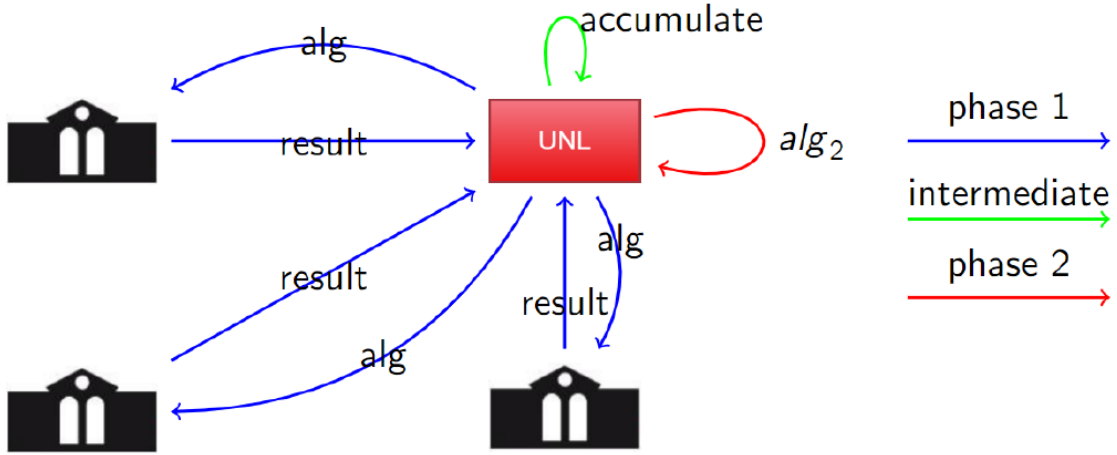
Many possible implementation techniques:
- Application: Policy Enforcement Point (PEP)
- Hardware: Trusted Execution Environments (TEE)
- OS/VM: Sandboxing
- Data: Cryptographic techniques (e.g. homomorphic encryption, key swapping)
- Middleware: e.g. Message interception
- 'Sticky Policies'

Popular use case: Digital Rights Management (DRM)

Commonalities:
You rely on the remote software/machine/infrastructure to enforce your policies on your behalf

Alternative: Compute-to-Data / Algorithm-to-Data

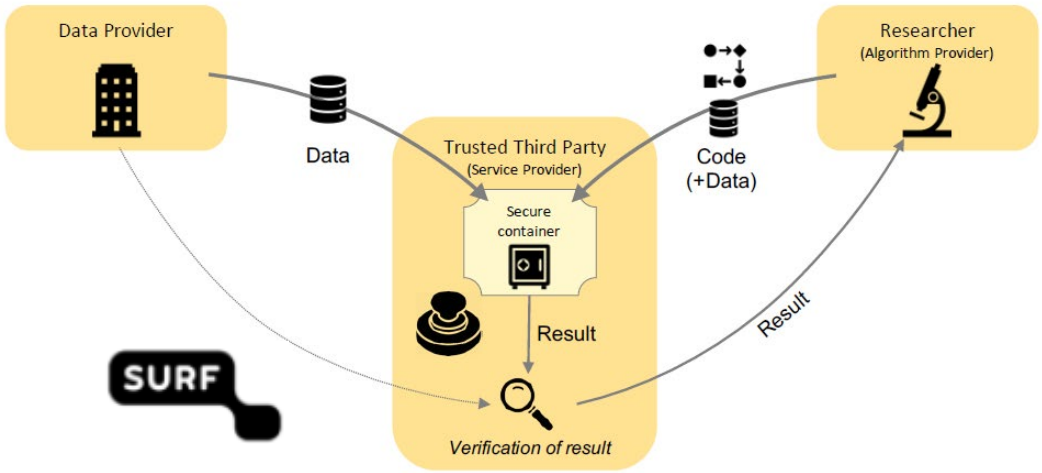
Keep the data secure and let the algorithm come to you (or to the data)



Multi-party "compute-to-data" archetype:

1. Participant UNL provides algorithm.
2. Participants run algorithm locally and share results.
3. UNL accumulates results and
4. runs second algorithm to yield insights

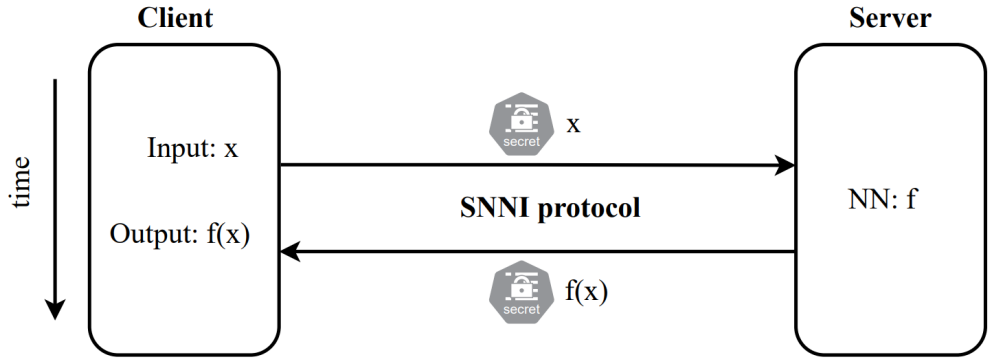
(For example: computing average salary across organisations)



In the "Sharing Data via Trusted Third Party (TTP)" archetype,

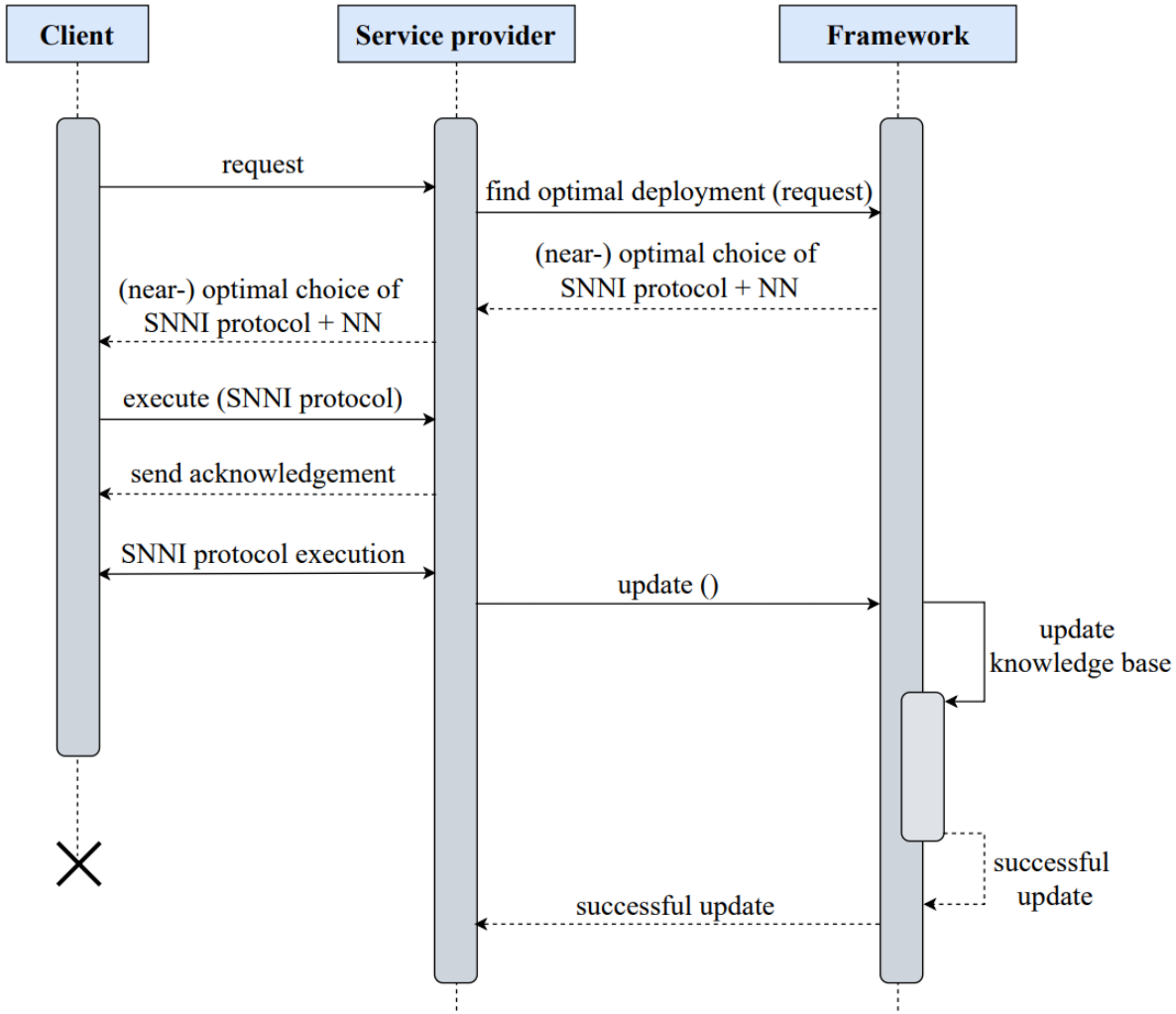
1. the Data Provider provides the data, and the Algorithm Provider provides the algorithm.
2. The algorithm is run in a secure container (e.g., 'no internet connection') at a Service Provider, and
3. the output of this computation is first verified by the data provider to ensure it does not contain any confidential information, only then is
4. the output is released to the Algorithm Provider.

Secure Neural Network Inference (-aaS)



- Context: Secure Inference as a service, securing the neural network to keep function `f` private whilst providing answers `f(x)`
- Goal: framework that assists the service provider (server) in selecting the most energy efficient Secure Inference techniques for the case

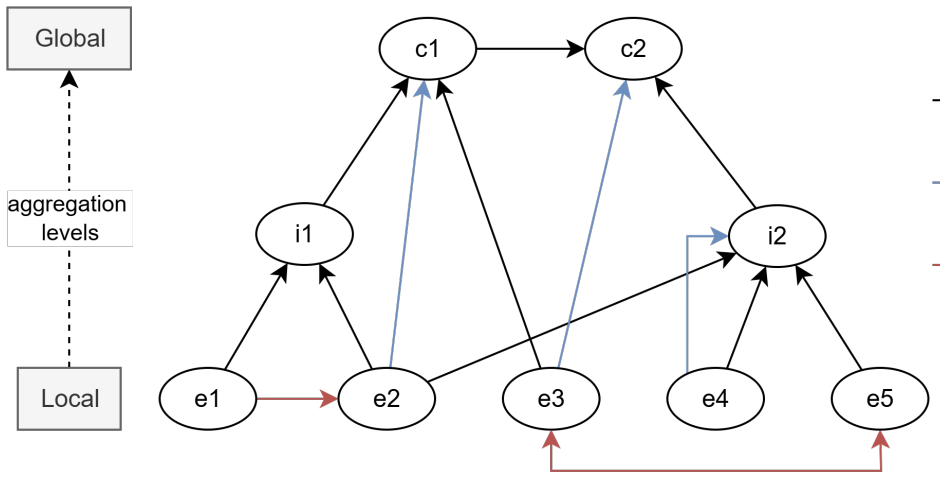
T. Islam, A. Oprescu, Z. Á. Mann and S. Klous, "A Framework to Optimize the Energy Cost of Securing Neural Network Inference,"

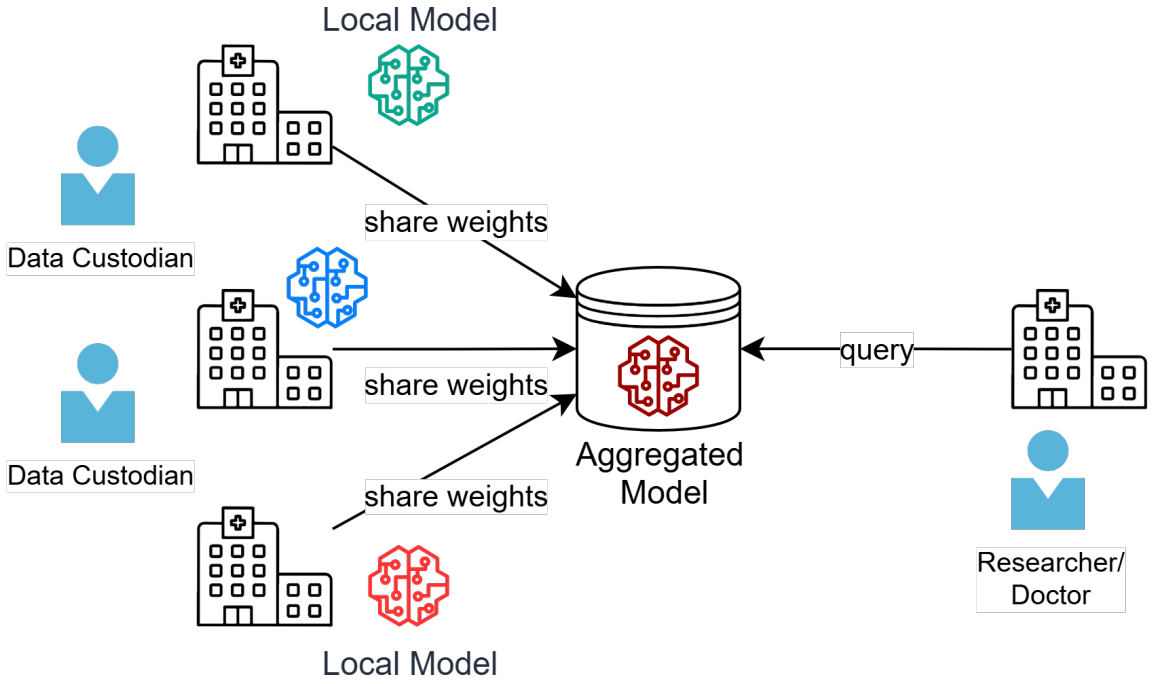


Federated Machine Learning

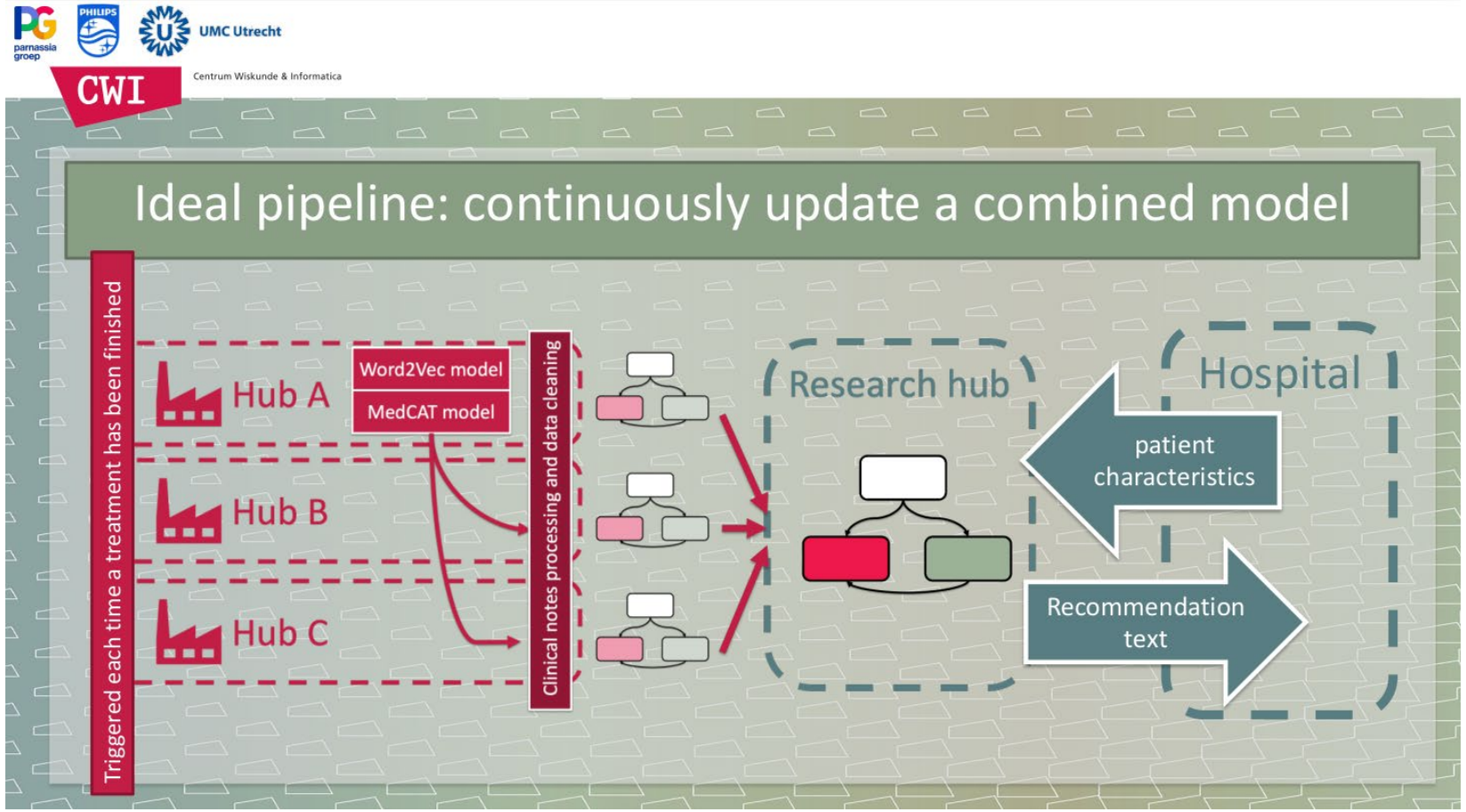
Organisations that are able and willing to perform computations on behalf of the consortium can apply a range of secure Multi-Party Computation (sMPC) techniques and Privacy Enhancing Techniques (PETs), such as:

- Federated Machine Learning
- Secret Sharing
- Data Synthesis
- Differential Privacy





sMPC (example case)



As part of the Enabling Personalized Interventions Project: https://enablingpersonalizedinterventions.nl/