Policy-driven distributed data processing

in the AMdEX data exchange architecture

L. Thomas van Binsbergen

ltvanbinsbergen@acm.org

Assistant Professor, Complex-Cyber Infrastructure, University of Amsterdam

With: Cees de Laat, Leon Gommans, Paola Grosso, Sander Klous, Tom van Engers, Wouter Los, and Christopher Esterhuyse, Milen Girma Kebede, Lu-Chi Liu, Mostafa Mohajeri Parizi, Merrick Oost-Rosengren

May 15, 2024

EFRD-funded: Amsterdam Data Exchange (AMdEX) Fieldlab

UNIVERSITEIT VAN AMSTERDAM

deXes

European Union
European Regional
Development Fund
Investing in your future

Regulated data exchange:

Data exchange systems governed by regulations, agreements and policies

as an instance of

Regulated systems:

software systems with embedded regulatory services derived from legal/regulatory specifications that monitor and/or enforce compliance

Regulated data exchange:

Data exchange systems governed by regulations, agreements and policies

as an instance of

Regulated systems:

software systems with embedded regulatory services derived from legal/regulatory specifications that monitor and/or enforce compliance

Requirement analysis

- Goal: systems with legally justifiable data exchange actions (sharing, processing)
- Solution ingredients: high-level specification, enforcement strategies, access and usage control, static and runtime verification

Section 1

Policy-driven data exchange @ UvA

Policy Administration and Enforcement

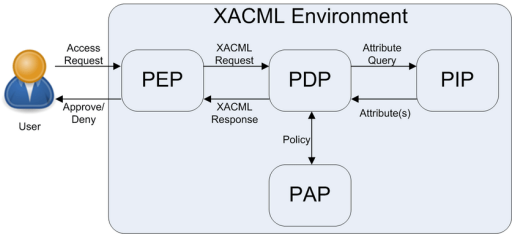


Figure: Simplified XACML architecture – M.S. Ferdous. "User-controlled identity management systems using mobile device". PhD thesis.

Requirements on Administration

- Links between legal text and policy
- Versioning, persistence
- Layered policies, level of abstraction
- Policy reuse, reusable templates
- Usability: registration, selection, ...

Policy Administration and Enforcement

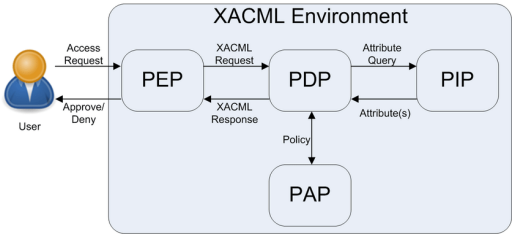


Figure: Simplified XACML architecture – M.S. Ferdous. "User-controlled identity management systems using mobile device". PhD thesis.

Requirements on Administration

- Links between legal text and policy
- Versioning, persistence
- Layered policies, level of abstraction
- Policy reuse, reusable templates
- Usability: registration, selection, ...

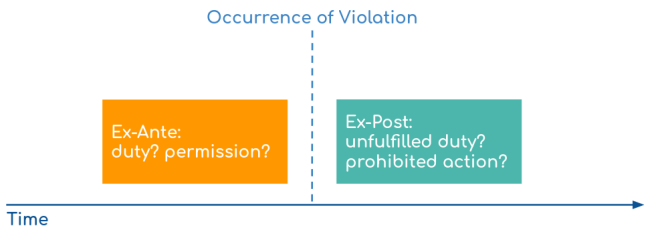
Requirements on Policy Language

- Connects legal primitives and computational primitives
- Compositional and extensible specifications
- Supports authorisation, scenario checking, simulation, verification

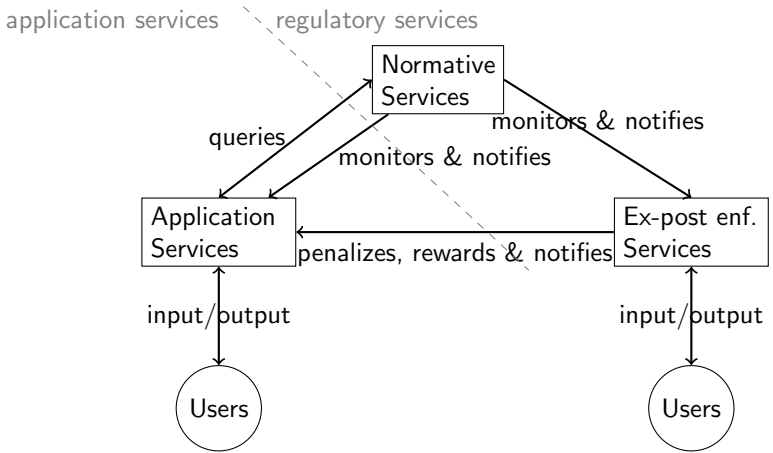
Policy Administration and Enforcement

Requirements on Enforcement

- Occurs at all stages:
  "before, during and after processing"
- Ex-ante and ex-post enforcement
- Legal obligations
- Accountable
- Explainable
- Pre- and post-conditions
- Human-in-the-loop



Regulated systems with ex-post enforcement



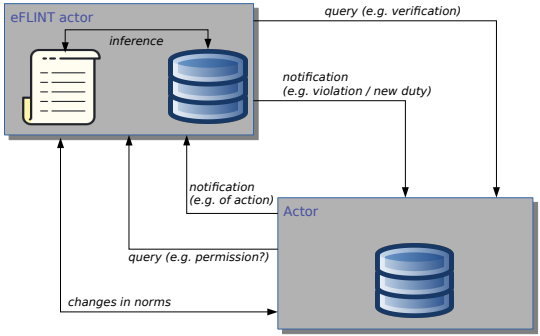
User interactions:

- Making observations triggering violations
- Confirming violations
- Acting on violations

Lu-Chi Liu, Mostafa Mohajeri Parizi, L. Thomas van Binsbergen, and Tom M. van Engers. "Regulatory Services to Automate Compliance with Ex-post Enforcement". In: *Proceedings of AICOL 2023*. 2024

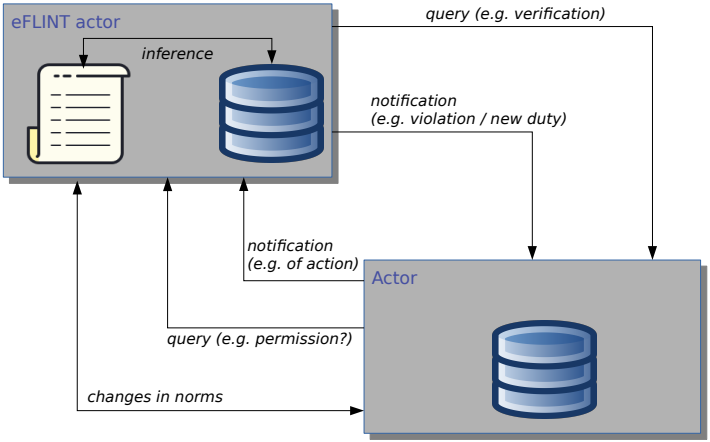
Policy reasoning with eFLINT domain-specific language (DSL)

Formalization of laws and policies
- *declarative reasoning about compliance: facts, actions and duties*
- *reactive service for software integration*
- *satisfies aforementioned requirements*
- *can be used to generate ODRL rules*



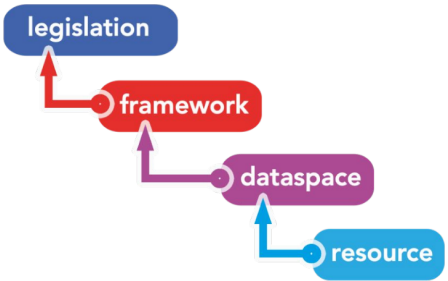
L. Thomas van Binsbergen, Lu-Chi Liu, Robert van Doesburg, and Tom M. van Engers. "eFLINT: a domain-specific language for executable norm specifications". In: *Proceedings of the 19th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*. ACM, 2020, pp. 124–136. DOI: 10.1145/3425898.3426958

Policy reasoning with eFLINT domain-specific language (DSL)



L. Thomas van Binsbergen, Lu-Chi Liu, Robert van Doesburg, and Tom M. van Engers. "eFLINT: a domain-specific language for executable norm specifications". In: *Proceedings of the 19th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*. ACM, 2020, pp. 124–136. DOI: 10.1145/3425898.3426958

Layered policy specification



Rule of law, International, EU and local

Trust eco-system & governance principles for sharing data

Consortium agreements "how we share data"

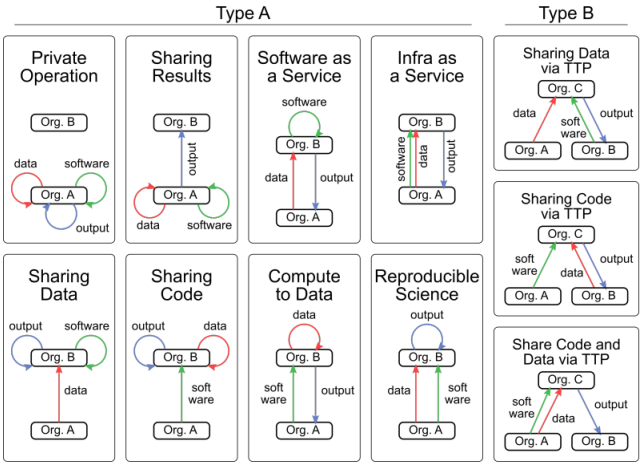
Conditions for sharing specific data, services, documents, applications

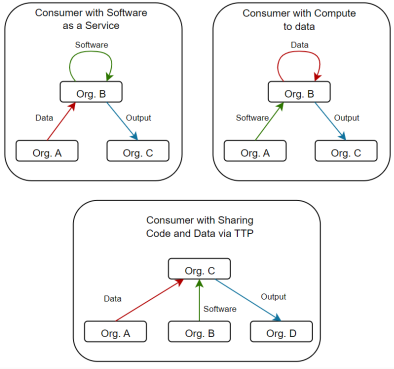
Experiments

- GDPR $\longrightarrow$ Financial sharing agreement $\rightarrow$ Organisational policy
- GDPR $\longrightarrow$ Medical consortium regulatory document $\rightarrow$ Resource-level access control

L. Thomas van Binsbergen, Milen G. Kebede, Joshua Baugh, Tom M. van Engers, and Dannis G. van Vuurden. "Dynamic generation of access control policies from social policies". In: *Proceedings of ICTH 2021*. Vol. 198. Procedia Computer Science. Elsevier, 2021, pp. 140–147. DOI: 10.1016/j.procs.2021.12.221

Reuse – Data exchange archetypes





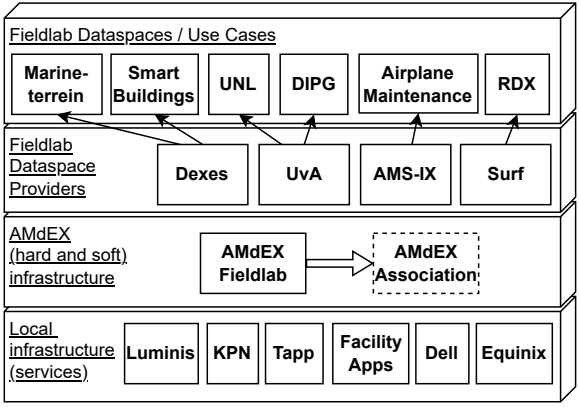
https://gitlab.com/eflint/data-exchange-templates (Nina Verheijen)

Sara Shakeri, Lourens Veen, and Paola Grosso. "Evaluation of Container Overlays for Secure Data Sharing". In: *2020 IEEE 45th LCN Symposium on Emerging Topics in Networking (LCN Symposium)*. 2020, pp. 99–108. DOI: 10.1109/LCNSymposium50271.2020.9363266

Section 2

AMdEX fieldlab

AMdEX fieldlab overview



L. Thomas van Binsbergen, Merrick Oost-Rosengren, Hayo Schreijer, Freek Dijkstra, and Taco van Dijk. *AMdEX Reference Architecture – version 1.0.0*. Ed. by L. Thomas van Binsbergen. Feb. 2024. DOI: 10.5281/zenodo.10565915

AMdEX Reference Architecture – roles

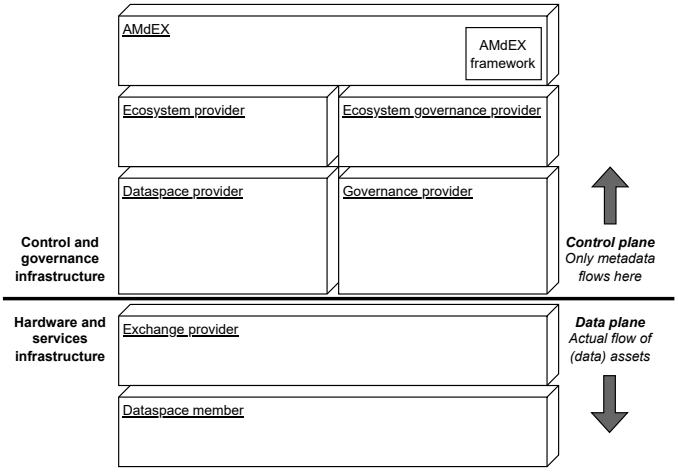


Figure: Infrastructural roles

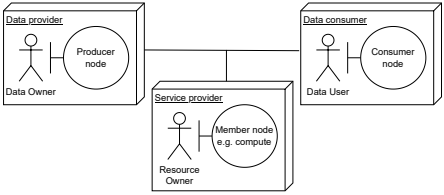


Figure: Dataspace member roles

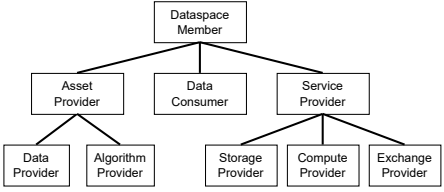


Figure: Member taxonomy

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**
3. **Offering**: members offer their resources through the **Catalog** and **Policy Store**

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**
3. **Offering**: members offer their resources through the **Catalog** and **Policy Store**
4. **Requesting**: a member proposes an application by submitting a workflow

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**
3. **Offering**: members offer their resources through the **Catalog** and **Policy Store**
4. **Requesting**: a member proposes an application by submitting a workflow
5. **Clearing**: authorizations gathered for workflow actions (**Enforcement Orchestrator**)

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**
3. **Offering**: members offer their resources through the **Catalog** and **Policy Store**
4. **Requesting**: a member proposes an application by submitting a workflow
5. **Clearing**: authorizations gathered for workflow actions (**Enforcement Orchestrator**)
6. **Processing**: workflow actions are executed and logged (**Process Orchestrator**)

Lifetime of data exchange applications

1. **Onboarding**: members get registered and connected via the **Registry**
2. **Proposing**: a member proposes consortium agreement, registered in the **Policy Store**
3. **Offering**: members offer their resources through the **Catalog** and **Policy Store**
4. **Requesting**: a member proposes an application by submitting a workflow
5. **Clearing**: authorizations gathered for workflow actions (**Enforcement Orchestrator**)
6. **Processing**: workflow actions are executed and logged (**Process Orchestrator**)
7. **Auditing**: logs are analysed for compliance (**Notary**),
   new information can be brought in (**Auditor**)

AMdEX fieldlab – main results

Main results and insights

- High-level reference architecture, software services at varying TRLs
- Main selling points: genericity (archetypes), integrated governance, legal requirements
- We have identified some important trade-offs:
  - Data privacy and sensitivity versus analytical power
  - Decentralized control versus accountability
  - Auditing requires access to several types of sensitive information

AMdEX fieldlab – main results

Main results and insights

- High-level reference architecture, software services at varying TRLs
- Main selling points: genericity (archetypes), integrated governance, legal requirements
- We have identified some important trade-offs:
  - Data privacy and sensitivity versus analytical power
  - Decentralized control versus accountability
  - Auditing requires access to several types of sensitive information

Next steps

- Consolidation and standardisation, interoperability with EU initiatives, i.e., IDSA and iShare
- **AMdEX-DMI** project: higher TRLs, research into partially automating auditing
- **Targeted use cases** with specific service providers:
  synthetic data, secure multi-party computation, federated ML, differential privacy, ...

Some open questions

- How general is our approach? How realistic is it to support generic archetypes?
  Can we sufficiently standardize to include many types of service providers?
  Howto secure multi-party computation (sMPC) and federated machine learning (FML)?

Some open questions

- How general is our approach? How realistic is it to support generic archetypes?
  Can we sufficiently standardize to include many types of service providers?
  Howto secure multi-party computation (sMPC) and federated machine learning (FML)?

- How realistic is our approach to policy administration and construction?
  Requires collaboration between legal and software expert?
  Many interpretations and versions across layers, how to prevent inconsistencies?

Some open questions

- How general is our approach? How realistic is it to support generic archetypes?
  Can we sufficiently standardize to include many types of service providers?
  Howto secure multi-party computation (sMPC) and federated machine learning (FML)?

- How realistic is our approach to policy administration and construction?
  Requires collaboration between legal and software expert?
  Many interpretations and versions across layers, how to prevent inconsistencies?

AMdEX-DMI project supported by the National Growthfund

DMI·ECOSYSTEEM

- How to trace and audit exchange processes when data, algorithms and logs are sensitive?

- What information is needed for auditing, and are service providers willing to share?
  Can we handle logging information as 'just another' sensitive data asset?
  Can we identify 'levels of auditability' to be recorded in agreements?

Policy-driven distributed data processing

in the AMdEX data exchange architecture

L. Thomas van Binsbergen

ltvanbinsbergen@acm.org

Assistant Professor, Complex-Cyber Infrastructure, University of Amsterdam

With: Cees de Laat, Leon Gommans, Paola Grosso, Sander Klous, Tom van Engers, Wouter Los, and Christopher Esterhuyse, Milen Girma Kebede, Lu-Chi Liu, Mostafa Mohajeri Parizi, Merrick Oost-Rosengren

May 15, 2024

EFRD-funded: AMDEX Fieldlab – neutral data-exchange infrastructure

Amsterdam Economic Board · SURF · amsix · Universiteit van Amsterdam · deXes · European Union, European Regional Development Fund, Investing in your future · Provincie Noord-Holland