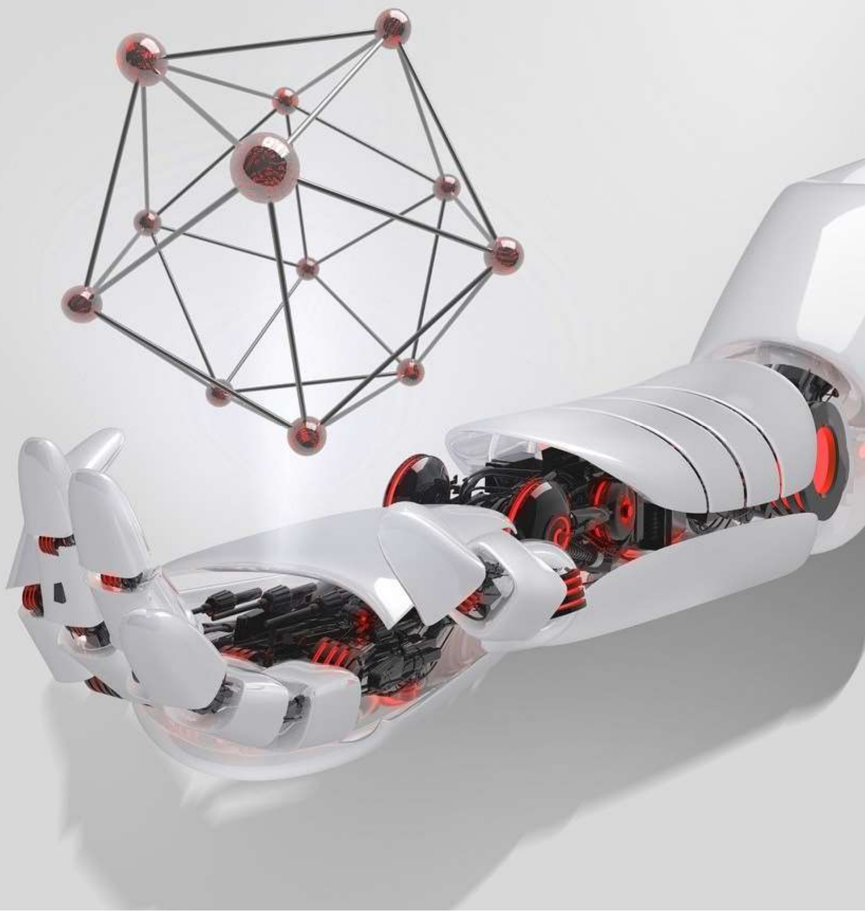
Data Centric Cyber Security

By Murat Kantarcioglu



1. Why Data Centric Cybersecurity and Privacy?

Data Centric Cybersecurity improves privacy, security, resiliency in more regulated environment and data-driven organizations

Value of Data Keeps Increasing

- BIG Data became the crown jewel of any organization
  - Customer data
  - Intellectual property

- Unlocking the value of data via machine learning and data science
  - Data gives new insights about the company
  - New products
  - More innovation

Changing Cybersecurity Landscape

- Data became an important target:
  - Cyber attacks against data
    - 63% increase in ransomware attacks 2023 2nd quarter **
- Data moves across the company and cloud services
- Remote work: Mobile devices and access to the data remotely
  - Cannot lock the organization.
- Need to protect data everywhere.
  - Zero-trust

** https://cyberint.com/blog/research/ransomware-trends-q2-2023-report/

Regulatory Compliance: Privacy Regulations

- Compliance with regulations
  E.g., GDPR
- Regulations require securing personally identifiable data in sharing and processing
- Leakage of data due to cyber attacks may trigger notifications.
- Cybersecurity policies may need to be integrated with privacy policies

GDPR security outcomes

This guidance describes a set of technical security outcomes that are considered to represent appropriate measures under the GDPR.

Source: iStock

" Manage security risk
Protect personal data against cyber attack
Detect security events
Minimise the impact" ***

*** Image Credit: https://www.ncsc.gov.uk/guidance/gdpr-security-outcomes

Data Centric Cybersecurity vs Traditional Cyber Security*

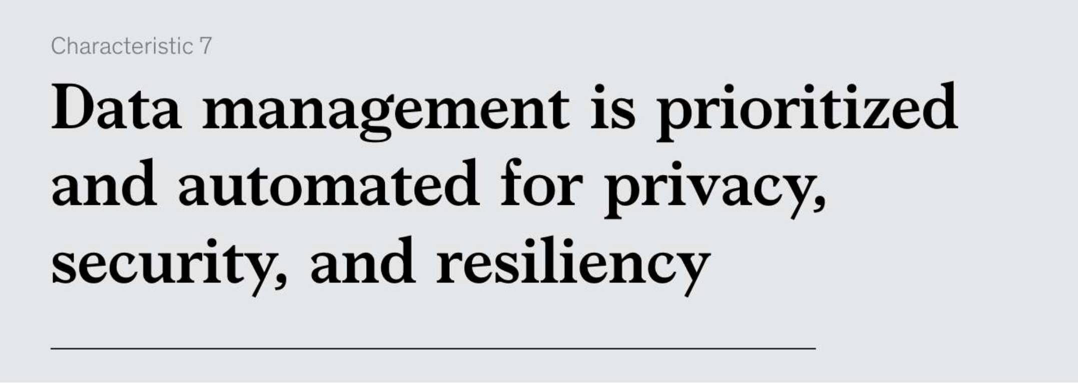
	Traditional Cyber Security	Data Centric Cyber Security
Data Visibility	LOW	**HIGH**
Data Governance	Almost Non-existent	**Required**
Perimeter Defense	YES	**YES and adds more layers**
Trust surface	HiGH	**LOW (Suitable for zero trust**
Ease of Compliance with Data Privacy Regulations	LOW	**HIGH**

- **Data visibility** quantifies whether cybersecurity mechanism is aware of the data type, location and sensitivity.
- “**Data governance** means setting internal standards—data policies—that apply to how data is gathered, stored, processed, and disposed of” **
- **Perimeter Defense**: Security protection such as firewall protecting the company from outside attacks
- **Trust surface**: The systems that are assumed to be trusted.

*“Data-centric security: Integrating data privacy and data”. Hennessy et al. , IBM Research Journal, Volume 53, No. 2.

** Definition taken from Google.

Business Impact of Data Centric Cybersecurity



Characteristic 7

Data management is prioritized and automated for privacy, security, and resiliency

"By 2025 **Organizational mindsets have fully shifted toward treating data privacy, ethics, and security as areas of required competency,** driven by evolving regulatory expectations such as the Virginia Consumer Data Protection Act (VCDPA), General Data Protection Regulation (GDPR), and California Consumer Privacy Act (CCPA); increasing consumer awareness of their data rights; and the increasingly high stakes of security incidents.

--The data-driven enterprise of 2025

- Protect one of the most important asset of any organization
  - Data !!
- Enable tracking of data lifecycle
  - Potential impact on data quality
- Easier compliance with privacy regulations
  - Implementing data centric cyber security makes it easier to comply with privacy regulations

***Image Credit: MCKINSEY: Data Driven Enterprise: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025

2. Pillars of Data Centric Cybersecurity

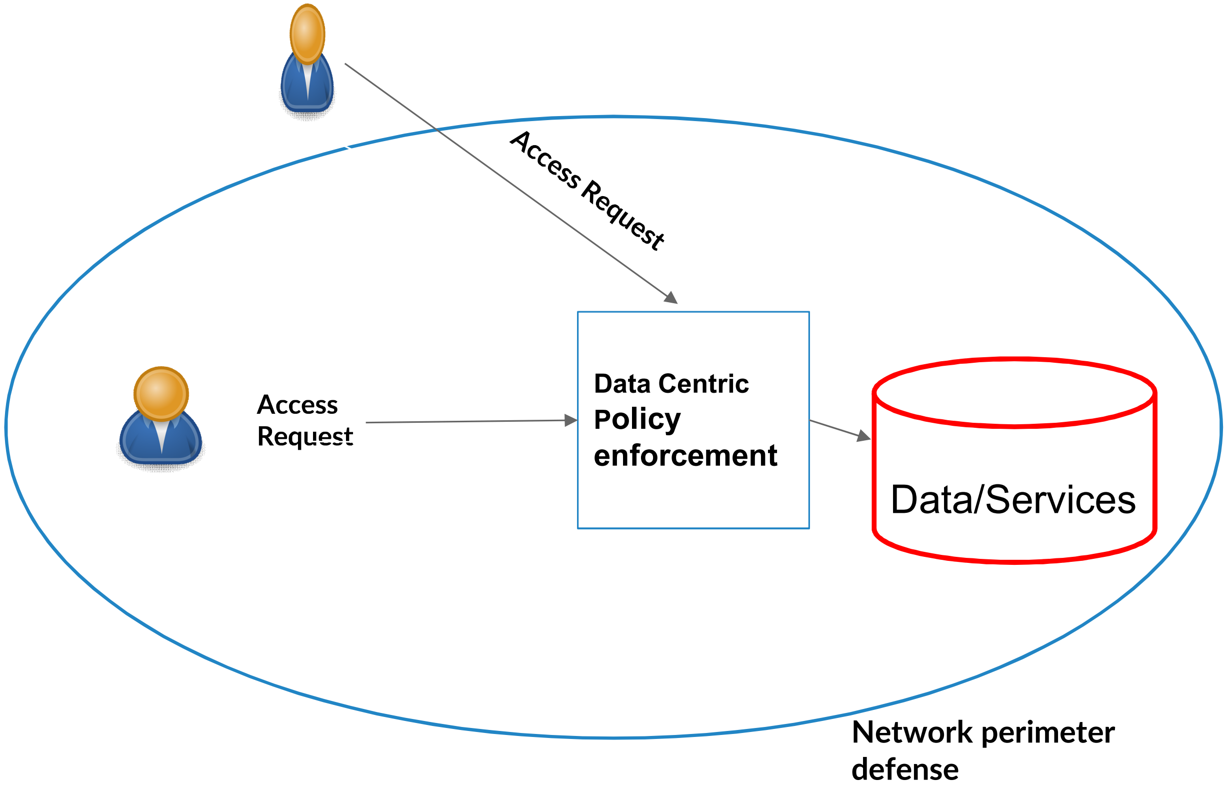
Major pillars of Data Centric Cybersecurity supports advanced data protection mechanisms.

Important Pillars

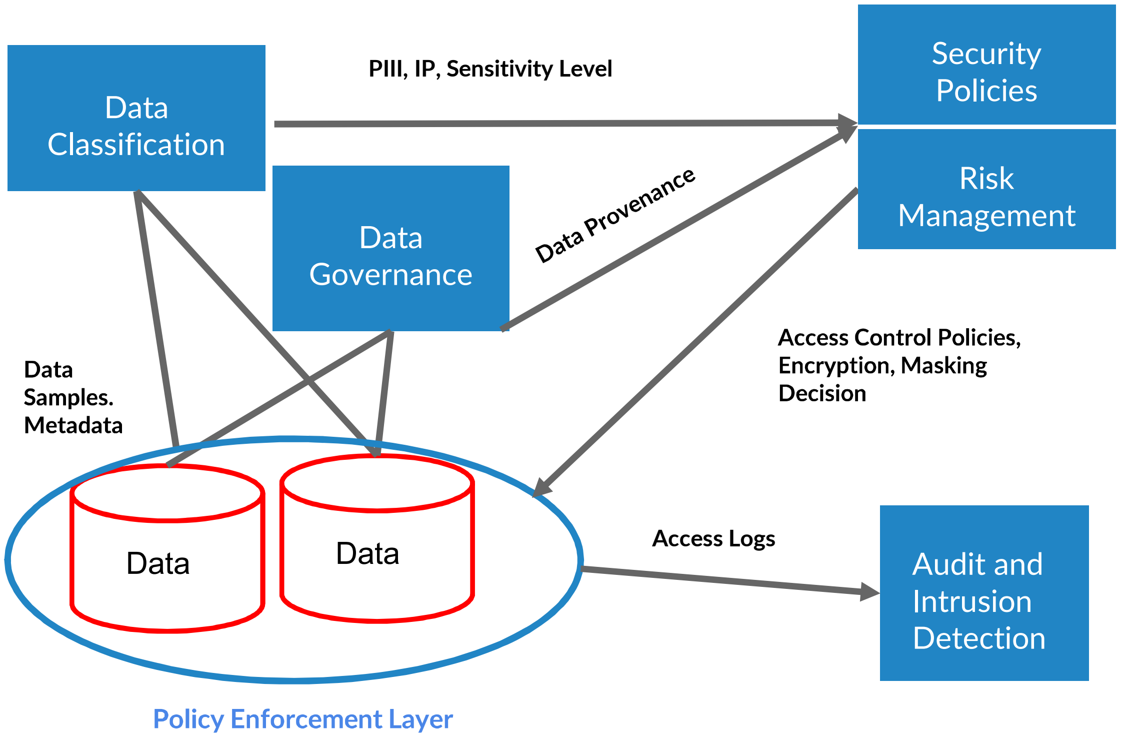
- Data Classification
  - Using AI to classify data
  - Tag data with appropriate labels
- Data Governance
  - Track the lifecycle of the data
  - Understand how the data is used and shared
- Security Policies
  - Define who can access to data and what they can do
  - Policies could be defined based on the attributes of the data, tasks and regulations

https://microsoft.github.io/presidio/

Data centric cyber security: Overview



Data centric cyber security architecture – 5 Pillars



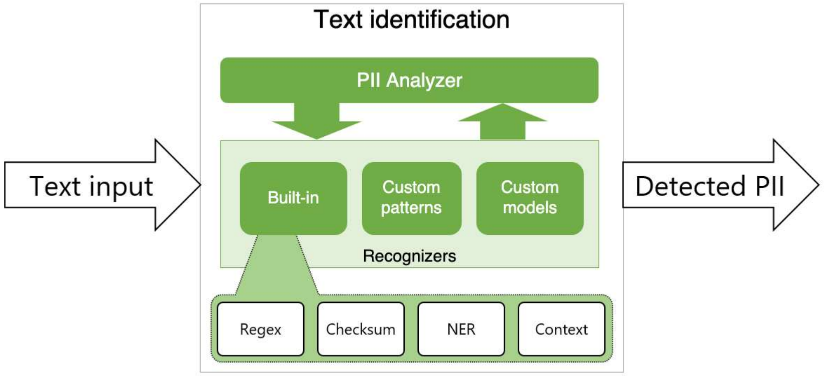
- **Blue boxes are basic pillars of the architecture.** These pillars are implemented as modules.
- **Arrows represent information flow**. For example, data samples go to Data Classification model for automatic data classification. Data classification sends results to Policy management module

Pillar 1: Data Classification – Automatic classification

- **Definition:** Classify data into data types.
- **Objective:** Cyber risks and compliance requirements depend on the data type so data types need to be understood
  - Automatically Classify DATA
    - Personally Identifiable Information
      - Emails
      - National Identifiers
      - Name
      - Surname
    - Intellectual Property
      - Sensitive product information

AI/ML/LLM models for data classification

Example: Microsoft Presidio



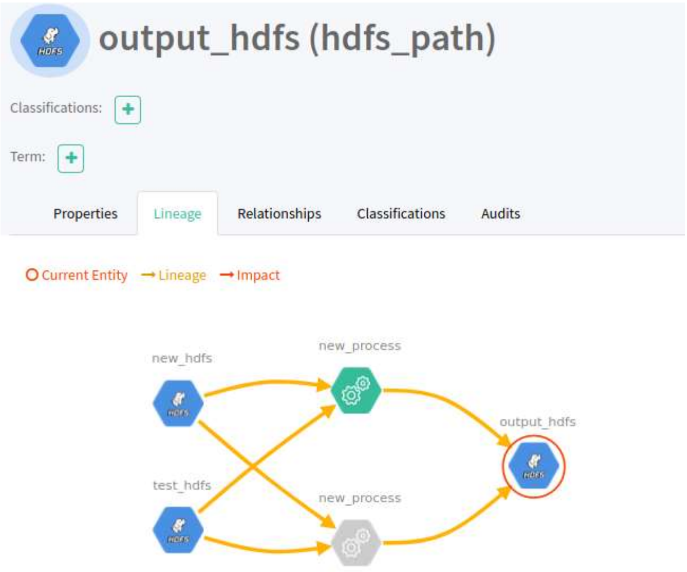
https://microsoft.github.io/presidio/

Pillar 2: Data Governance – following the flow of data

- **Definition:** “Data governance means setting internal standards—data policies—that apply to how data is gathered, stored, processed, and disposed of”**

- **Objective:** Need to follow the flow of the data across the organization
  - Most organizations do not know where all the data is and who accesses the data
  - Data location, access to sensitive data, processing of the sensitive data needs to be tracked.
  - Data may need to be deleted for compliance reasons or reducing potential attack surface

Example: Apache Atlas

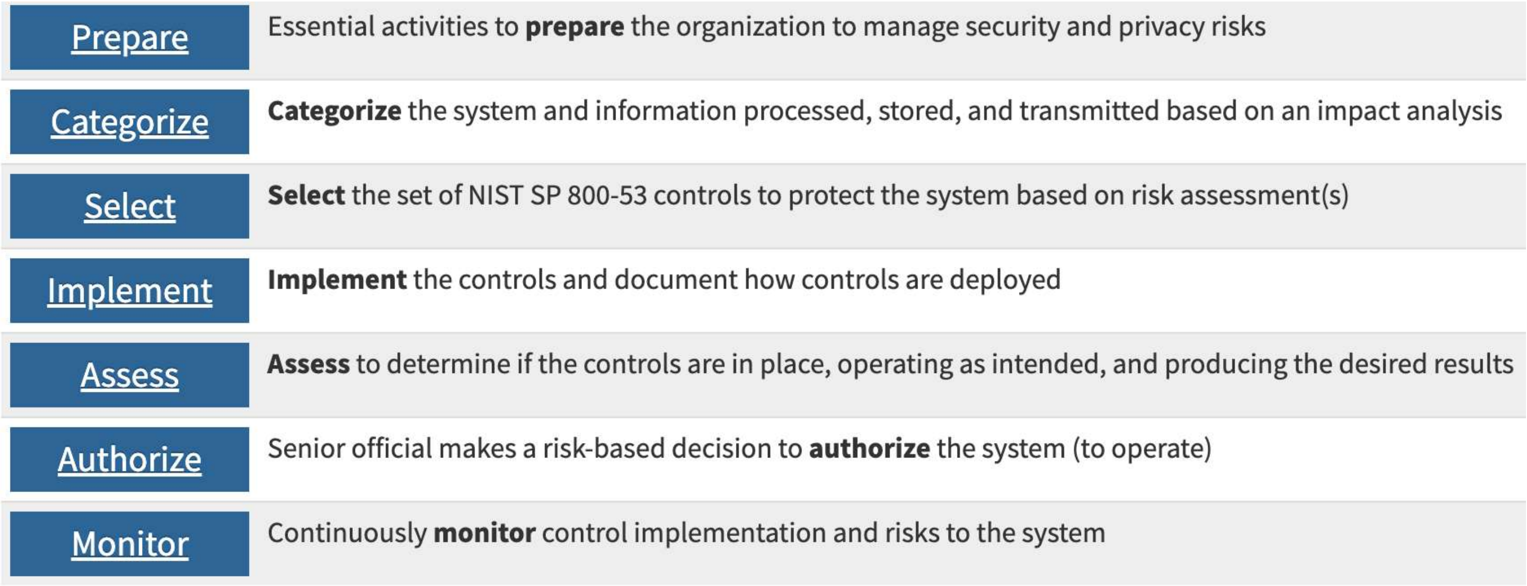


** Definition from Google Inc.

Pillar 3: Risk Management – Understanding and Reducing Cyber Risks

- **Definition:** Manage Cyber Risks
- **Objective:** Understand the cyber risks and take actions to reduce it.
  - Understanding cyber risks:
    - Cyber attacks:
      - Data Leakage
      - Ransomware
    - Regulatory compliance
    - Accidental data deletion
  - Reducing risks:
    - Security Policy Enforcement (More on this)
    - Data Sanitization
    - Data Deletion
    - Data Encryption at Rest and transit

Pillar 3: Risk Management Example: NIST Framework**



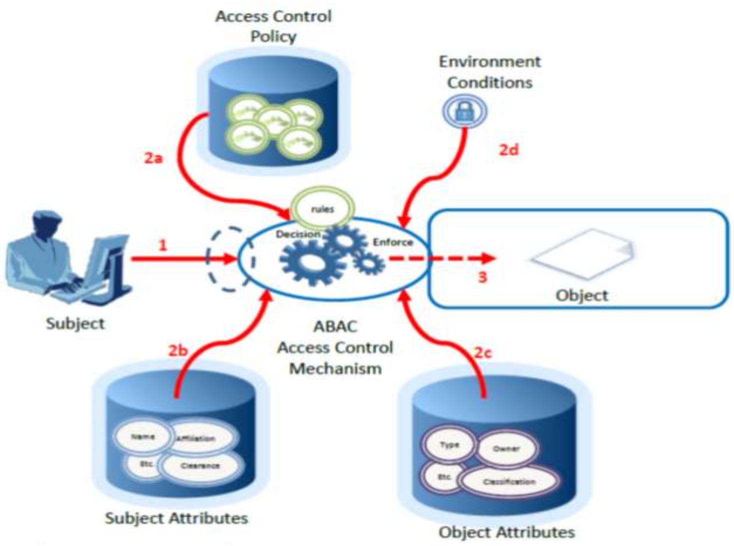
Prepare	Essential activities to **prepare** the organization to manage security and privacy risks
Categorize	**Categorize** the system and information processed, stored, and transmitted based on an impact analysis
Select	**Select** the set of NIST SP 800-53 controls to protect the system based on risk assessment(s)
Implement	**Implement** the controls and document how controls are deployed
Assess	**Assess** to determine if the controls are in place, operating as intended, and producing the desired results
Authorize	Senior official makes a risk-based decision to **authorize** the system (to operate)
Monitor	Continuously **monitor** control implementation and risks to the system

** Image credit: https://csrc.nist.gov/projects/risk-management/about-rmf

Pillar 4: Security Policies - Important part of Zero trust and Data Centric Solutions

- **Definition:** Access to data and resources based on the organizational policies
- **Objective:** Limit risks by restricting access to critical data
  - E.g., No super admin account that can access everything
  - Security Policies:
    - Important part of Zero trust and Data Centric Solutions
    - Policies may be defined based on
      - Users and Applications' Roles
      - Attributes of
        - Data
        - Users
        - Processes
        - Location
        - Context

Example: NIST **



1. Subject requests access to object
2. Access Control Mechanism evaluates a) Rules, b) Subject Attributes, c) Object Attributes, and d) Environment Conditions to compute a decision
3. Subject is given access to object if authorized

**Image taken from NIST ABAC Standard

Pillar 5: Sensitive Data - Intrusion Detection

- **Definition:** Detects unauthorized or malicious access to data and/or resources.
- **Objective:** Machine Learning based Anomaly Detection
  - Check for predefined policies and historical access patterns to detect anomalous behaviour

3. Applications of Data Centric Cybersecurity

Data Centric Cybersecurity provides advanced technology, features, and applications

Data Centric Cyber Security Providers

Some features of existing products

- Advanced data governance
- Advanced data sanitization
- Complementary to existing security systems
- Access Control

BlueTalon protegrity

IMMUTA Informatica

DATAGUISE datiphy™

Federated Data Access: Potential Business Impact

- **Goal:** Sharing data for achieving business goals
  - E.g., optimize supply chains across multiple organization's
  - Potential to unlock huge value by combining data across
  - organizations
- **Objective**: Security and Privacy needs to be addressed
  - Sharing the entire data sets is too risky
  - Sharing only what is need is critical
  - Policy enforcement needed on each data owner’s organization

Federated Data Access: Example: IDSA

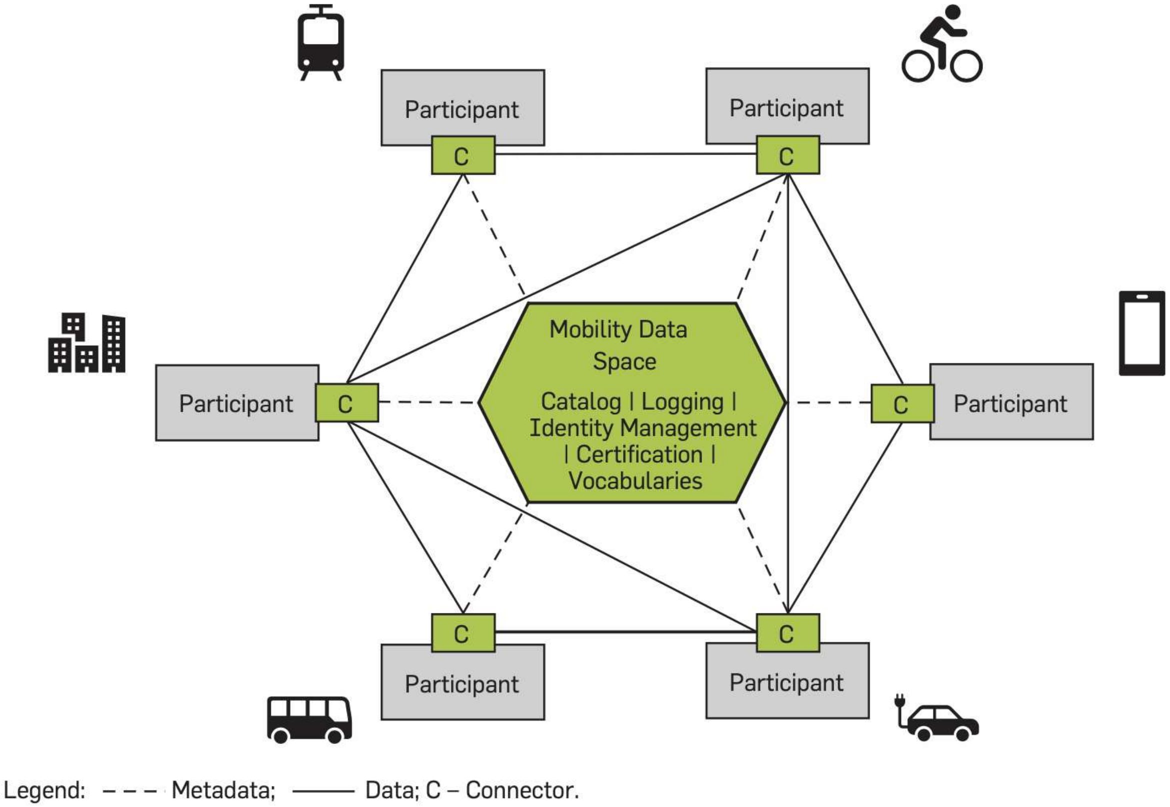
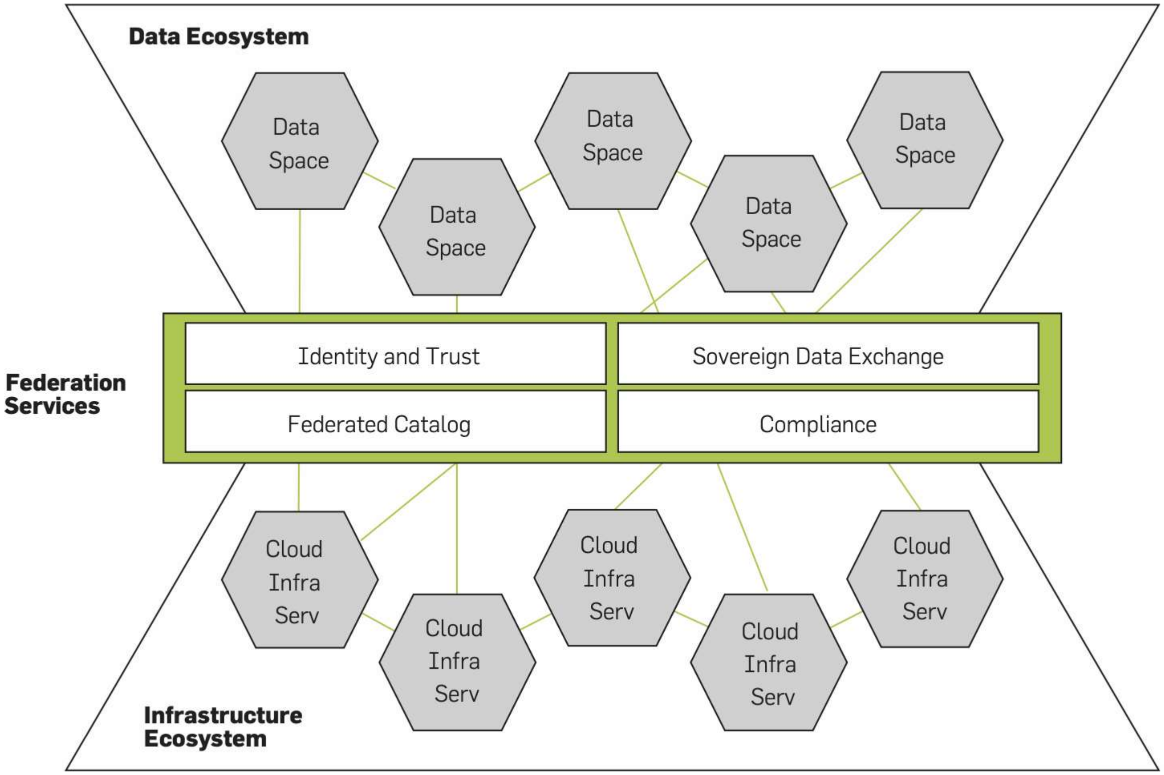

Participant
C
Participant
C
Participant
C
Mobility Data Space
Catalog | Logging | Identity Management | Certification | Vocabularies
C
Participant
C
Participant
C
Participant
Legend: – – – Metadata; —— Data; C – Connector.

Image from DOI:10.1145/3512341

Federated Data: Example: Gaia-x**



Data Ecosystem

Data Space

Data Space

Data Space

Data Space

Data Space

Federation Services

Identity and Trust

Sovereign Data Exchange

Federated Catalog

Compliance

Cloud Infra Serv

Cloud Infra Serv

Cloud Infra Serv

Cloud Infra Serv

Cloud Infra Serv

Infrastructure Ecosystem

Image from DOI:10.1145/3512341

4. Future of Data Centric Cybersecurity

Industry standards are emerging, integrating best practices remains challenging, tools and technologies are evolving.

Current Limitations and Challenges

- Access Control and Data Governance technologies are well understood and developed.
- Integrating best practices and tools **for access control, data governance and data centric intrusion detection** across the company infrastructure still remain a significant challenge.
  - Different data management systems with different access control and audit capabilities so it is very hard to implement security policies uniformly across different data management systems
  - Each system has its own logs. Organizations need to integrate the logs coming from different systems to get the accurate picture with respect to intrusions.
  - Usually different companies provide access control and data governance tools. Integrating these tools from different companies emerge as an important challenge.

Future Trends / Suggestions

- Sharing Data across Organizations:
    - I.e., Gaia-X type projects
    - Standards are still emerging
        - I.e. Catena-X for auto industry
- **Hardware based confidential computing could be influential for data sharing across organizations**
    - Using hardware based trusted execution environments may allow end to end encrypted data processing
    - Could enhance tools like AWS Cleanroom
- As AI/ML become a core part of any organization security of these AI/ML models would be critical.
    - Protecting the AI models could emerge as an important challenge.
    - Security policies related to AI model training and access needs to be considered.
    - Risk management related to AI security should be conducted. I.e., what happens if your AI model is attacked?

Future Trends

- Cryptographic tools like Homomorphic Encryption and Secure multi-party computation could prove to be revolutionary for data sharing across organizations
    - My personal belief is that they will not be cost effective enough in the near future for big data.
    - This technology is part of the risk management and could be used to reduce risks while federated data sharing
    - I believe it will take at least 10-15 years for these tools to gain significant traction.

- DCC will allow the better protection of the important company asset: DATA! And make compliance with regulations easier.

Thank you.

Questions?