Data Governance Act and GDPR in the scope of EMERALDS

An introductory overview of key EU regulations shaping data protection, privacy, and AI development

Datasets with Anonymized Data

- **Dataset 5 - Weather information**: Clearly, weather data doesn't contain personal information, hence inherently anonymized.
- **Dataset 8 - Bridge openings**: Mentioned as not containing personal data, implying it is anonymized regarding personal identification.
- **Dataset 9 - Incidents**: Specified as not containing personal data; therefore, it's anonymized.
- **Dataset 10 - Roadworks**: Also noted as not containing personal data, hence anonymized.
- **Dataset 13 - E-tickets validation data**: Mentioned that data is already anonymised by the provider.
- **Dataset 14 - Public Transport Route list data GTFS**: Data is stated to be anonymised by the provider.

Datasets Anonymized with a probability of PII exposure

- **Dataset 1 - Event Calendar Scheveningen**
- **Dataset 2 - Parking data**
- **Dataset 3 - Public transport data**
- **Dataset 4 - Shared Mobility**
- **Dataset 6 - Mobile App Counting**: Described as having closed access due to potential sensitivity but not explicitly mentioned as anonymized.
- **Dataset 7 - Bicycle counting Netherlands**
- **Dataset 11 - Floating Car Data**: Confidential, intended for consortium use, suggesting possible personal data handling but not explicitly stated as anonymized.
- **Dataset 12 - Loop Detector Data**: Similar to Dataset 11, confidential and for consortium use without explicit mention of anonymization.
- **Dataset 15 - Vehicle events data, GPS**: Indicated as confidential, shared within the consortium, and the text does not explicitly state the data is anonymized.
- **Dataset 16 - GPS Traffic Data for the North Yorkshire**: Mentioned as closed access, with no explicit mention of anonymization

Navigating EU Regulations for Responsible AI

GDPR expands individual privacy rights

GDPR gives EU citizens more control over their personal data and imposes obligations on organizations that process it.

Data Governance Act enables data sharing

The DGA aims to make more data available for reuse and facilitates data sharing between businesses and governments.

AI Act proposes risk-based regulations

The AI Act will introduce mandatory requirements for high-risk AI systems related to datasets, transparency, human oversight, etc.

Understanding how these regulations interact allows for developing ethical and legally compliant AI systems.

GDPR in the Context of AI

GDPR as a cornerstone of data protection laws in the EU

GDPR provides a comprehensive legal framework for data protection in the EU

Key principles of GDPR like consent, right to access

GDPR strengthens individual rights over personal data like consent requirements, right to access data

Rules for transferring data outside the EU

GDPR has strict requirements for transferring personal data outside the EU

GDPR establishes core data protection requirements in the EU relevant to development of AI systems

GDPR vs. AI Act: Distinguishing the Focus

GDPR focuses on data protection and privacy

The GDPR aims to give individuals more control over their personal data through principles like purpose limitation and data minimization.

AI Act takes a risk-based approach

The AI Act classifies AI systems based on their level of risk to regulate them accordingly, with higher-risk systems facing more scrutiny.

GDPR has wider applicability than AI Act

While the AI Act focuses specifically on AI, the GDPR's data protection principles apply more broadly to all personal data processing activities.

The GDPR and AI Act are complementary pieces of legislation, with the GDPR establishing baseline data protection standards and the AI Act adding specific AI governance mechanisms.

GDPR's data protection obligations

- **Data protection by design and default**

  GDPR requires data protection measures to be built into products and services from the earliest stages of development.

- **Cross-border data transfer rules**

  GDPR establishes conditions and safeguards for transferring personal data outside of the EU to ensure protection is maintained.

- **Data minimization**

  GDPR mandates that personal data collection and processing should be limited to what is directly relevant and necessary to accomplish a specified purpose.

- **Storage limitation**

  GDPR requires that personal data is only retained for as long as needed to fulfil the purposes for which it was collected.

- **Integrity and confidentiality**

  GDPR obligates entities to implement appropriate security measures to protect the integrity and confidentiality of personal data.

Integrating EMERALDS Datasets for AI: FAIR Principles at Work

Focus on Datasets 3 & 4

Public Transport and Shared Mobility datasets demonstrate integrating diverse mobility data to enrich AI inputs

Interoperability

Unified data models enable AI systems to efficiently process diverse datasets

Accessibility

API interfaces ensure accessibility of datasets in line with FAIR principles

Integrating diverse mobility datasets using unified models and APIs improves AI learning while ensuring FAIR principles.

Ensuring AI Ethics and GDPR Compliance

Ensure anonymization of sensitive mobility data

Use differential privacy techniques when handling datasets like vehicle GPS data to preserve privacy

Minimize data collection and retention

Only collect necessary data for the AI system's purpose and delete it when no longer needed to comply with data minimization principles

Conduct continuous ethics reviews

Review AI systems and data practices regularly to address emerging data protection concerns and maintain GDPR compliance

Implementing strong data governance practices through anonymization, minimization, and ethical reviews helps ensure AI systems using sensitive data remain compliant with GDPR and respect user privacy.

DGA's data sharing frameworks

The Data Governance Act (DGA) establishes frameworks to facilitate the sharing and reuse of certain categories of data across sectors and EU member states. This includes provisions enabling access to and reuse of public sector data under specific conditions. For projects developing AI and data analytics applications like GREEN.DAT.AI, the DGA's data sharing mechanisms can enable valuable data access to help drive innovation and societal benefits.

Data Governance Act: Beyond AI Regulation

Introduce Data Governance Act

Provide a brief introduction to the Data Governance Act, its aims and key concepts

Explain data sharing mechanisms

Elaborate on how DGA facilitates data sharing through data intermediaries and data altruism

Discuss impact on EU economy

Highlight the significance of DGA for innovation and economic growth in the EU

The Data Governance Act establishes a framework to enable secure and ethical data sharing in the EU, which can drive innovation and economic growth.

Unique Contributions of the Data Governance Act

Facilitating data sharing and reuse

The DGA establishes frameworks to enable easy and secure data sharing between sectors and EU Member States.

Supporting data intermediaries

The DGA provides guidance on roles and responsibilities for data intermediaries to enable a thriving data economy.

Encouraging data altruism

The DGA introduces new models like data altruism to incentivize voluntary data sharing for the common good.

The Data Governance Act introduces unique mechanisms to foster responsible data sharing and power innovation across the EU.

Data Governance Challenges with Sensitive Datasets

Mobile App Counting data anonymization

Closed access dataset with potential for future sharing raises questions on maintaining anonymity and consent when scaling or integrating data.

Vehicle GPS data anonymization

Confidential dataset with potential future openness requires careful handling of GPS data to prevent indirect identification of individuals.

Key data governance challenges include maintaining anonymity and consent when sharing sensitive datasets like Mobile App Counting and Vehicle GPS data.

DGA's data intermediaries and altruism

Data Intermediaries

The DGA establishes data intermediaries to enable secure, lawful data sharing between organizations.

Trusted Third Parties

Data intermediaries serve as trusted third parties facilitating data exchange and enhancing trust.

Data Altruism

The DGA enables voluntary data contributions from individuals/companies for purposes benefitting society.

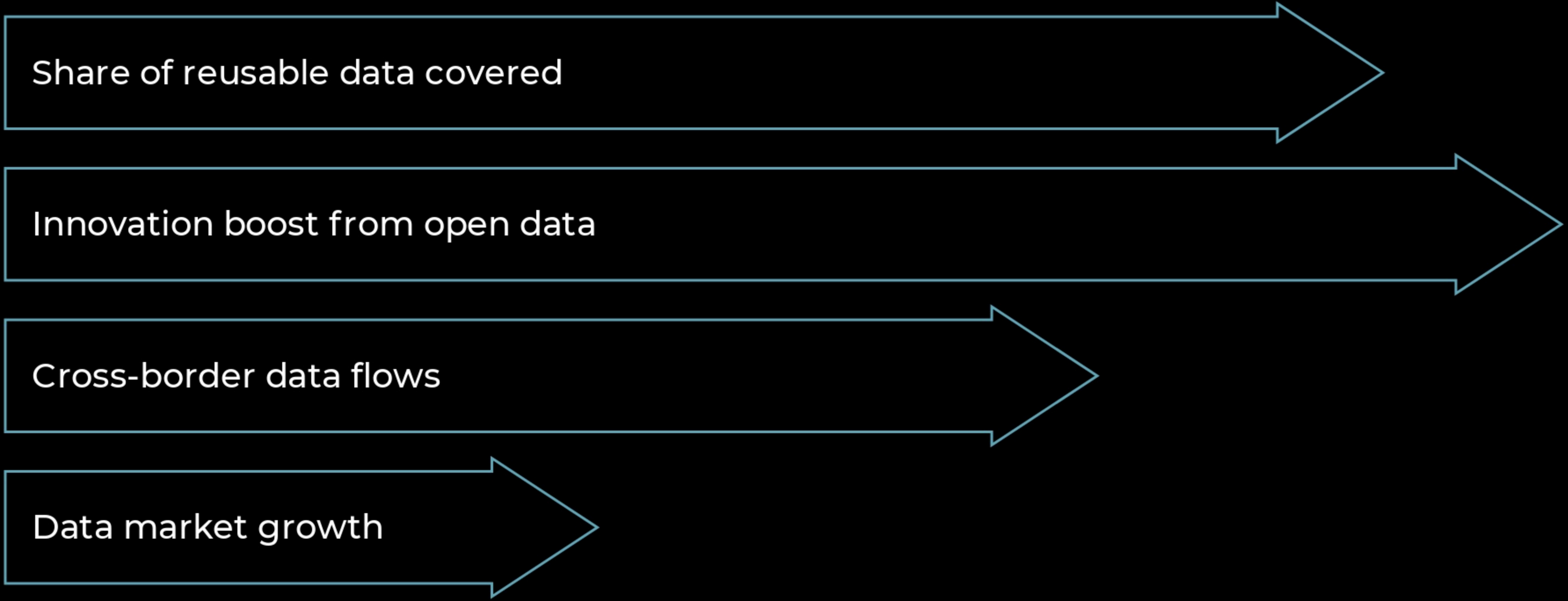
Informed Consent

Data altruism relies on freely given, specific, informed and unambiguous consent of data subjects.

Common European Data Spaces

Data intermediaries and altruism facilitate data use and reuse, contributing to common European data spaces.

DGA's data access and use



Share of reusable data covered

Innovation boost from open data

Cross-border data flows

Data market growth

Synergies and Distinctions

GDPR foundations

GDPR provides a baseline for data protection and privacy.

DGA builds further

DGA adds provisions for data access, sharing, reuse across sectors.

AI Act regulates AI systems

AI Act introduces requirements for trustworthy and ethical AI.

Together, these form a comprehensive EU regulatory environment for ethical and responsible data and AI innovation.

DGA Enables Responsible AI Development: The Scheveningen case

Scheveningen Event Calendar provides data on upcoming events

This data helps predict mobility demands during events

DGA mandates data quality checks

Ensures integrity of Event Calendar data used by AI systems

DGA enables secure cross-border data sharing

Allows AI to incorporate data from across EU for better predictions

DGA facilitates data use in AI for mobility prediction while maintaining quality and security

GDPR Compliance Strategies for Scheveningen Mobility Data

Sensitive nature of mobility data

Highlight how mobility data in Scheveningen dataset can reveal personal information like event attendance and movement patterns

Anonymization techniques

Use anonymization methods to remove personally identifiable information from the data

Minimize data collection

Only collect necessary data for the AI application to limit privacy risks

Obtain explicit consent

Get opt-in consent from individuals before collecting or processing their personal data

Implementing GDPR principles like anonymization and consent helps balance innovation and privacy when developing AI solutions using Scheveningen mobility data.

Anonymized Datasets Comply with GDPR, unless PII is exposed otherwise

Five datasets have anonymized data

Weather, bridge openings, incidents, roadworks and public transport data are anonymized by providers

Datasets comply with GDPR

Anonymization protects personal information as required by GDPR, unless identifiable otherwise

Many datasets follow best practices for data privacy and compliance.

Data Governance Act and GDPR Compliance: the case of not explicitly anonymised datasets

Anonymization

11 datasets lack explicit mention of anonymization, posing potential GDPR compliance risks.

Access Controls

Several datasets have closed/confidential access but unclear if sufficient for GDPR.

Personal Data

At least some datasets probably contain personal data

Recommendations

Review all datasets for PII risks, implement anonymization and access controls to ensure GDPR compliance.

Legal Guidance

Consult DPO's to confirm GDPR protections for datasets are adequate.

Training

Provide GDPR and data privacy training to teams managing datasets.

Conclusion: A Holistic Approach to AI and Data in the EU

Adhere to GDPR

Ensure AI and data projects comply with GDPR data protection principles.

Follow DGA

Make sure AI systems align with DGA requirements for high-risk AI systems.

Respect rights

Design AI to respect individual rights like non-discrimination and transparency.

By following GDPR, DGA, and the AI Act, we can build trustworthy AI that respects rights and manages risks.

Questions & Answers

1 | Q&A