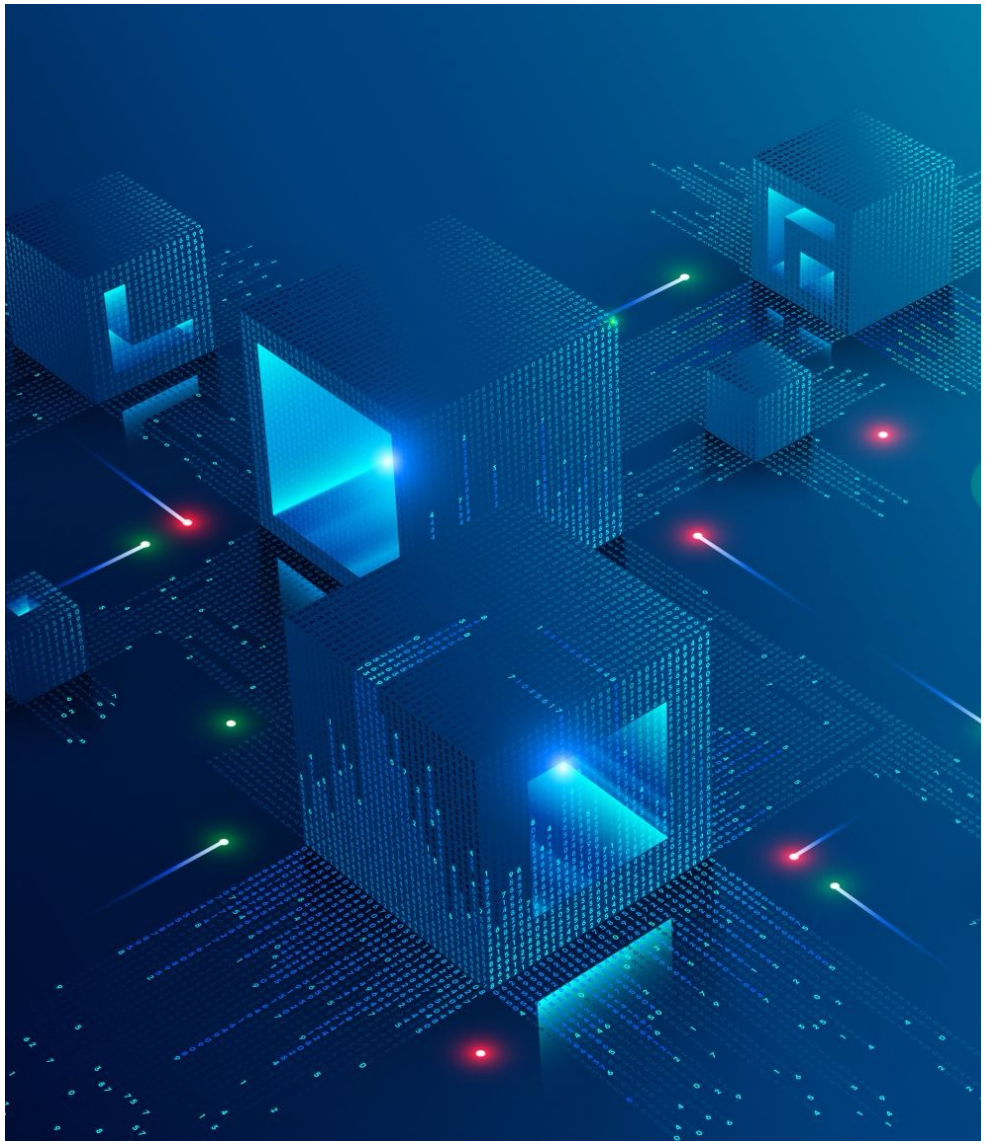
Data Governance: Organizing Data for Trustworthy Artificial Intelligence



Author & Context:

Written by Marijn Janssen, Paul Brous, Elsa Estevez, Luis S. Barbosa, Tomasz Janowski

The research paper focuses on:

The critical role of data governance in ensuring the trustworthiness of AI systems. It examines how data, the foundation of AI decision-making, must be managed responsibly to mitigate risks such as bias, lack of transparency, and ethical violations

Challenges in Data Governance for AI

Complexity: AI uses vast, high-velocity data from multiple sources, making governance difficult.

Bias & Ethics: Poor data governance can result in biased decisions, discrimination, and privacy violations.

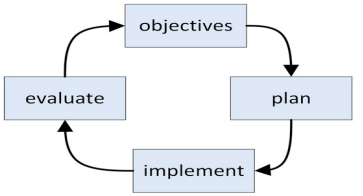
Security & Privacy: Data must be protected from cyber threats, unauthorized access, and misuse.

Regulatory Compliance: Organizations must adhere to GDPR and other legal requirements.

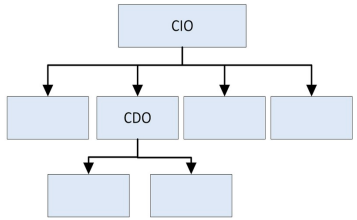
Approaches

The paper identifies three approaches to data governance:

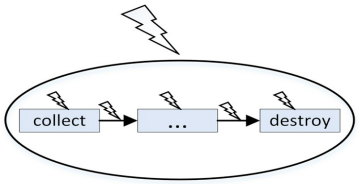
1. **Planning & Control Approach**: Establishes policies and structured monitoring to ensure data quality.
2. **Organizational Approach**: Assigns roles like Chief Data Officers (CDOs) to oversee AI governance.
3. **Risk-Based Approach**: Focuses on identifying AI-related risks (e.g., bias, security threats) and mitigating them.



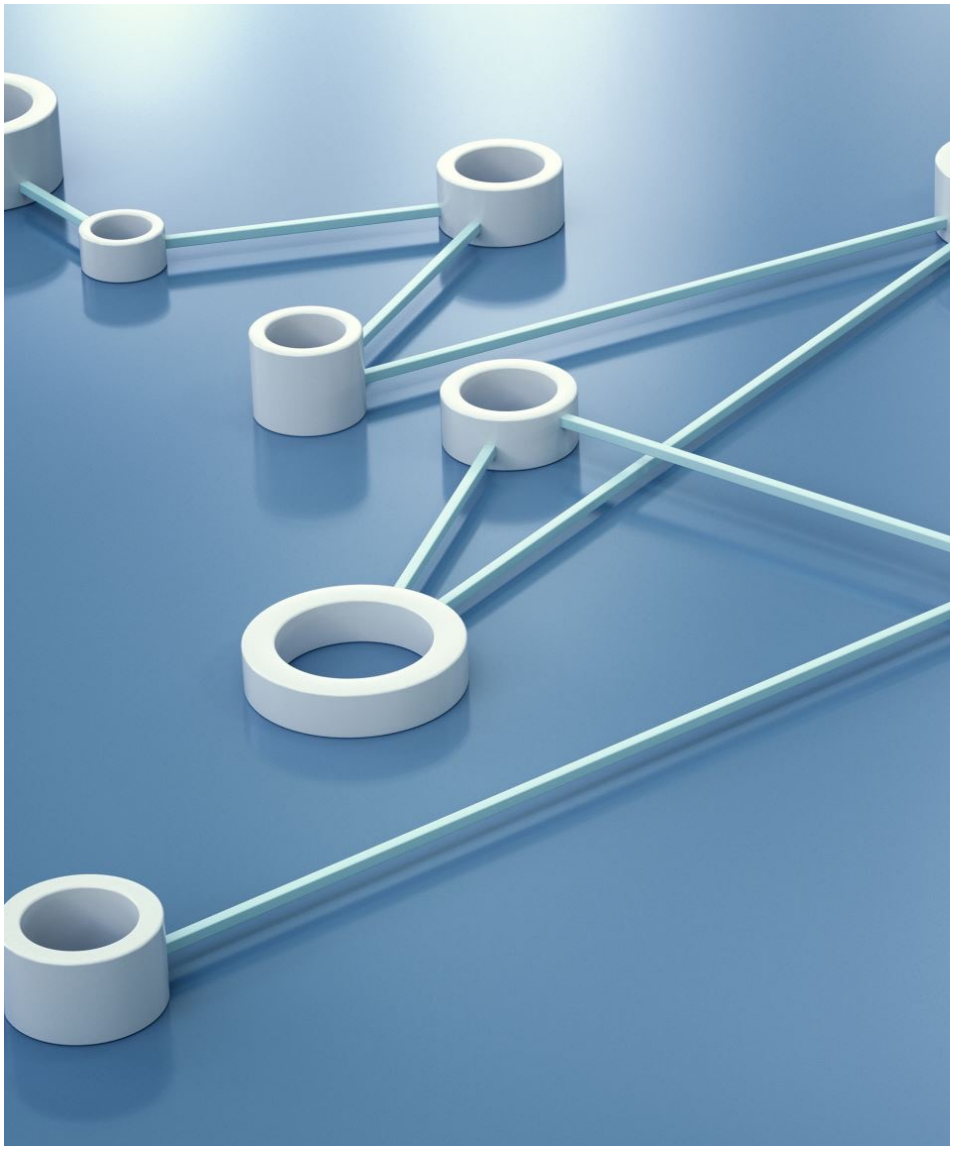
1. Planning and control



2. Organizational



3. Risk-based



Trusted Data Sharing Framework

- **Data Security Principles**: Ensures that only authorized entities access and use data.
- **Blockchain & Distributed Ledger Technology (DLT)**: Provides tamper-proof records of data transactions.
- **Self-Sovereign Identity (SSI)**: Allows users to control their own data.
- **Non-Repudiation Mechanisms**: Ensures transparency and data authenticity.

Key Principles of Data Governance

- **Evaluating Data Quality & Bias**: Ensuring datasets are accurate, unbiased, and suitable for AI applications.
- **Detecting Changing Patterns**: Monitoring AI-driven decisions to detect anomalies and inconsistencies.
- **Need to Know Principle**: Limiting data access to authorized personnel only.
- **Bug Bounty Programs**: Encouraging ethical hackers to detect security vulnerabilities in AI models.
- **Transparency in Data Sharing**: Ensuring all data exchanges are auditable and communicated with stakeholders.
- **Data Stewardship & Accountability**: Assigning roles to oversee responsible data usage.
- **Algorithmic Explainability**: Ensuring AI decisions can be interpreted and justified.

Summary

- **AI systems must have structured governance to avoid risks of bias, privacy breaches, and unethical decision-making.**
- **Organizations should implement a combination of planning, risk-based, and organizational approaches to data governance.**
- **Trusted data-sharing mechanisms (blockchain, encryption, and identity verification) can enhance AI trustworthiness.**
- **Regular audits, explainability in AI, and algorithmic fairness checks are essential for responsible AI governance.**

Citizen-centered data governance in the smart city: From ethics to accountability

Author & Context:

Written by Pascal D. König, published in *Sustainable Cities and Society* (2021).

The paper looks at how data in smart cities are managed, the ethical issues involved and accountability.

Introduction

- **Significance / Importance**

The paper explores how using data to create value in smart cities creates challenges for decision-makers and how they can handle these challenges.

- **Goals / Objectives**

The framework turns ethical ideas into clear governance mechanisms while ensuring they align with democratic accountability and citizen centric governance.

- **Relevance to class subject matter**

The paper aligns with data governance and ethics by addressing challenges in smart city governance, accountability, and citizen rights.

It discusses framework to balance innovation with ethical concerns.

Major Discussion Topics

1. Ethical and Legitimacy Challenges in Smart City Data Governance

- **Concept:**

The governance of smart cities involves ethical risks such as privacy violations, biased algorithms, and lack of transparency.

- **Example:**

The case of smart meters collecting energy consumption data illustrates privacy risks and the importance of data minimization.

2. The Data Value Chain and Its Implications

- **Concept**:
  - The paper categorizes data processes into three stages: collection, processing, and use.
  - Ethical issues emerge at each stage, requiring tailored governance mechanisms.
- **Example**:
  - The Hong Kong Octopus Card system, which sells citizen movement data to third parties.
- **Relevance**:
  - Highlights how unchecked data sharing can lead to unintended ethical violations.

The Hong Kong Octopus Card system

- The city of Hong Kong, introduced the Octopus Card, which can be used to pay for services like transportation and parking and as a key for accessing buildings.
- Through this card, the city has collected personal information about users and their behaviours.
- This information was not only used for managing city operations but had also been sold to third parties.

3. Accountability Framework for Smart Cities

- **Concept**:

  The paper proposes an accountability-driven governance model.

  Introduces three accountability mechanisms:

  - **Ex-Ante Testing**: Stakeholder consultations and impact assessments before implementation.
  - **Operational Transparency**: Public access to data governance processes and algorithmic explainability.
  - **Ex-Post Scrutiny**: Regular audits and reviews to ensure ongoing compliance.

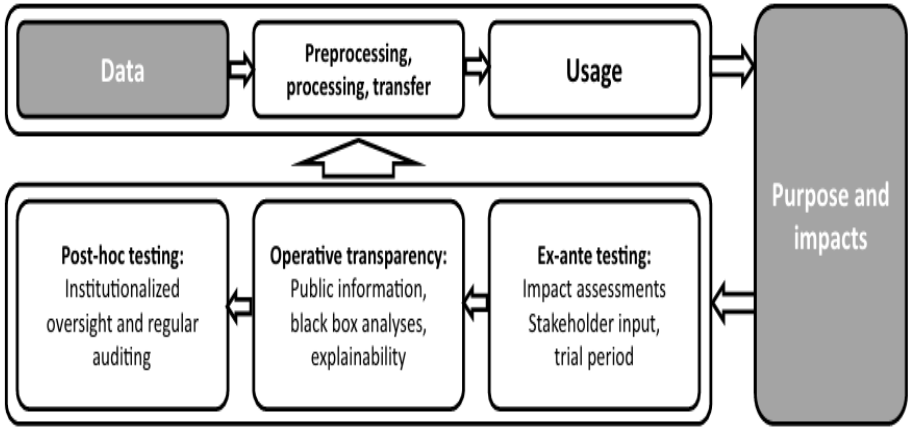


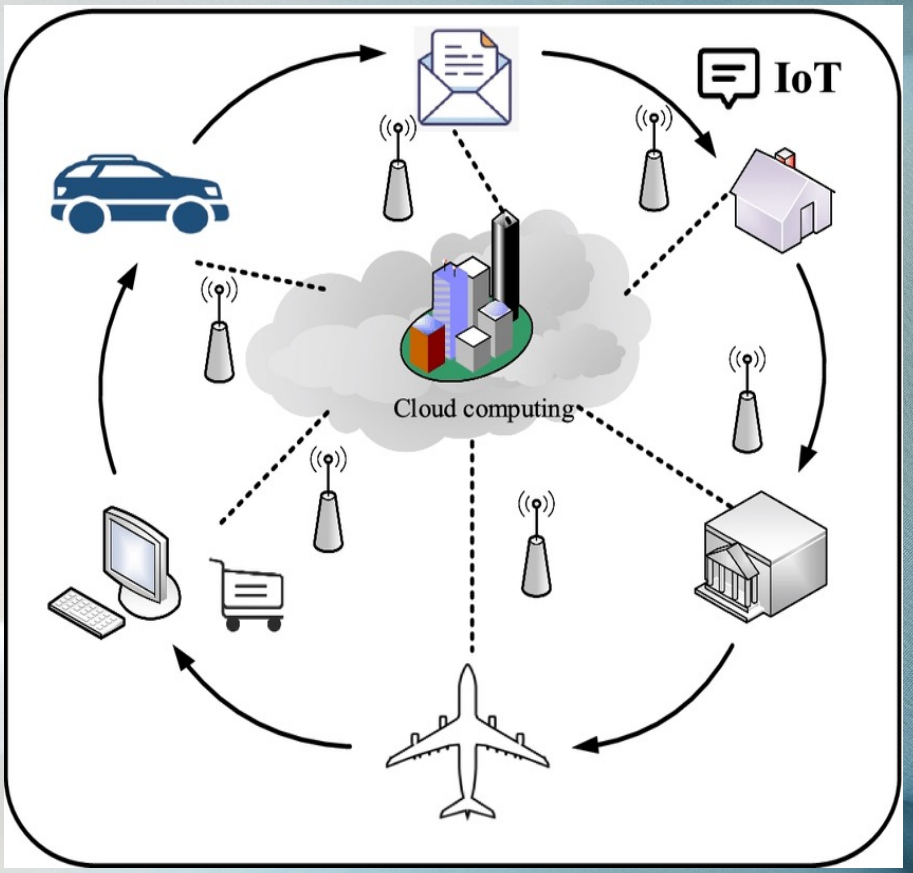
Fig. 2. Accountability for Citizen-Centered Data Governance.

Conclusion

- The paper provides a well-structured argument for a citizen-centered approach to data governance.
- It bridges ethical theories with practical governance frameworks.

- **Author's Critical Approach**:
  - The paper points out that current smart city management systems lack accountability.
  - Supports governance structures that are open and accountable while staying flexible with new technology.

Digital-Twin Technology for Smart & Sustainable tourism



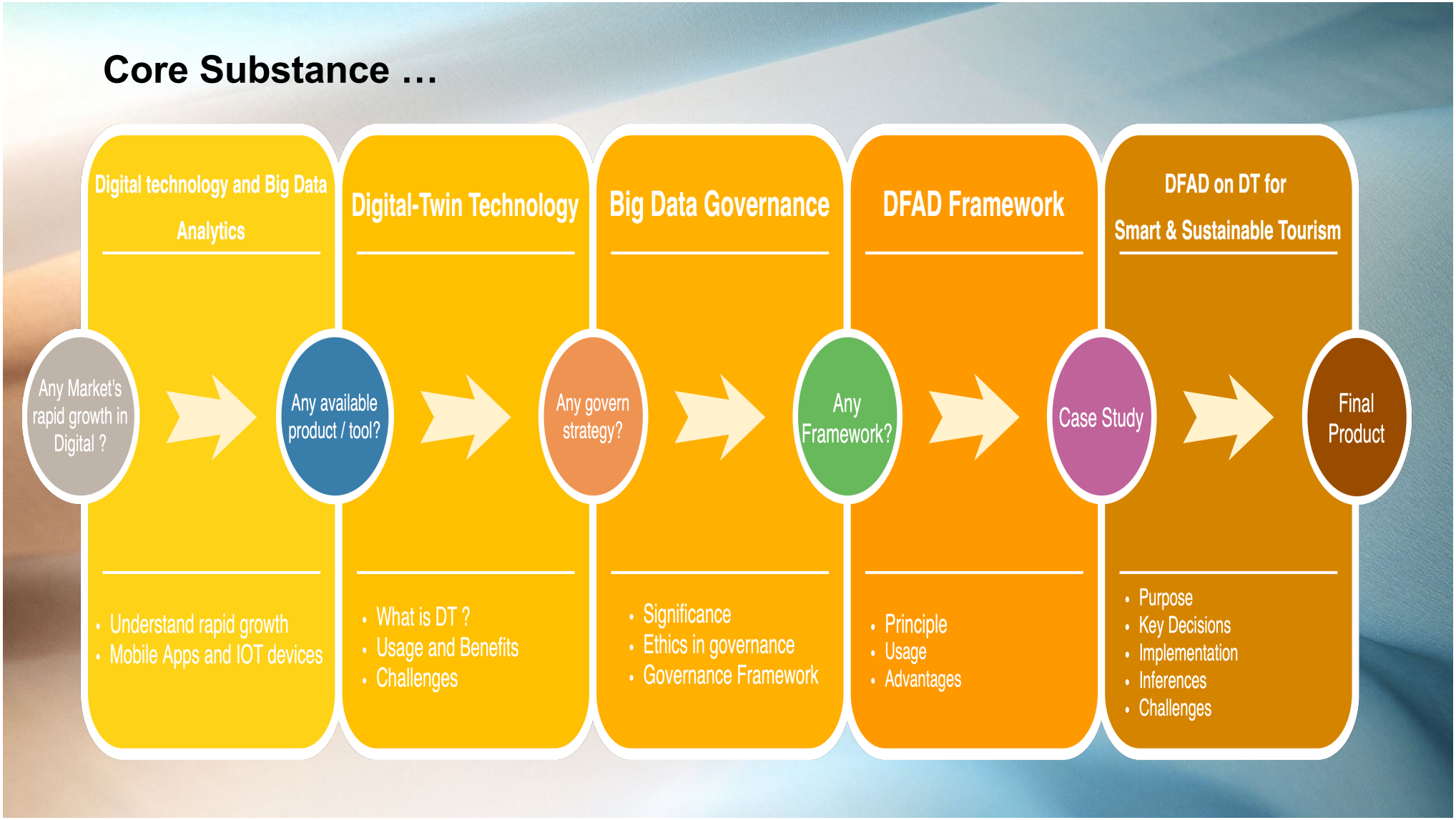
Author & Context

1. Eko Rahmadian
2. Daniel Feitosa
3. Yulia Virantina3

Accepted: 13 October 2023 / Published online: 16 November 2023

The paper proposes a Documentation Framework for Architectural Decisions (DFAD) to enhance big data governance in Digital Twin (DT) technology for smart and sustainable tourism, using Mobile Positioning Data (MPD) in Indonesia while ensuring transparency, security, and regulatory compliance

Core Substance …



Digital Technology with Big Data

Rapid growth in Digital technology with Big Data as a key driver for Innovation and Progress for sustainable development, furthermore integrated with AI, ML, and IoT

Leveraging Big Data Analytics - predict travel demand, make informed decisions, improved customer interactions and services, personalised market campaigns

Tourism mobile applications utilize big data analytics to provide visitors information, enabling analysis of travel routes and their impact on the environment

IoT based systems for personalized navigation in smart museums, improvising visitor's cultural experience

Digital-Twin Technology

1. Virtual Model that leverages real-time and historical data from IoT, AI, and Big Data resources
2. Enable predictive analysis, better decision-making for stakeholders
3. Implemented in many sectors and industries, including smart cities leading to Smart & Sustainable Tourism
4. With the rapid growth of real-time and historical data as strategic asset, effective data management and governance has become a key concern for organizations

Big Data Governance

Significance

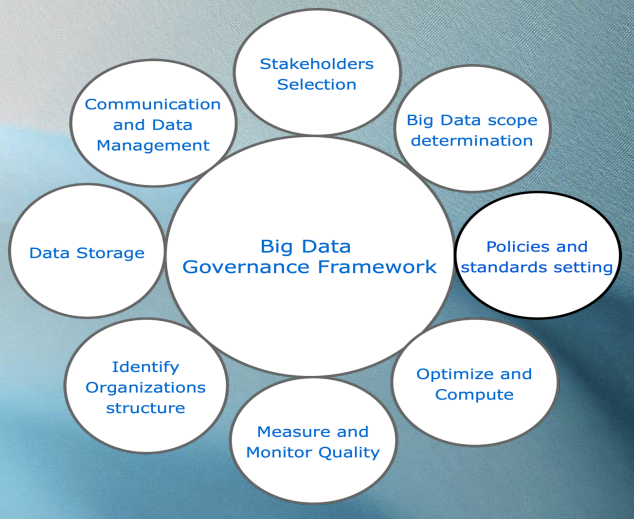
- Crucial in this digital realm to create legitimacy
- Ensures high data quality, meet compliance laws and regulations, particularly those related to privacy and security
- Helps organisations use data effectively for decision making and business success
- Yang et al. (2019) warn that business disruptions increase data breach risks

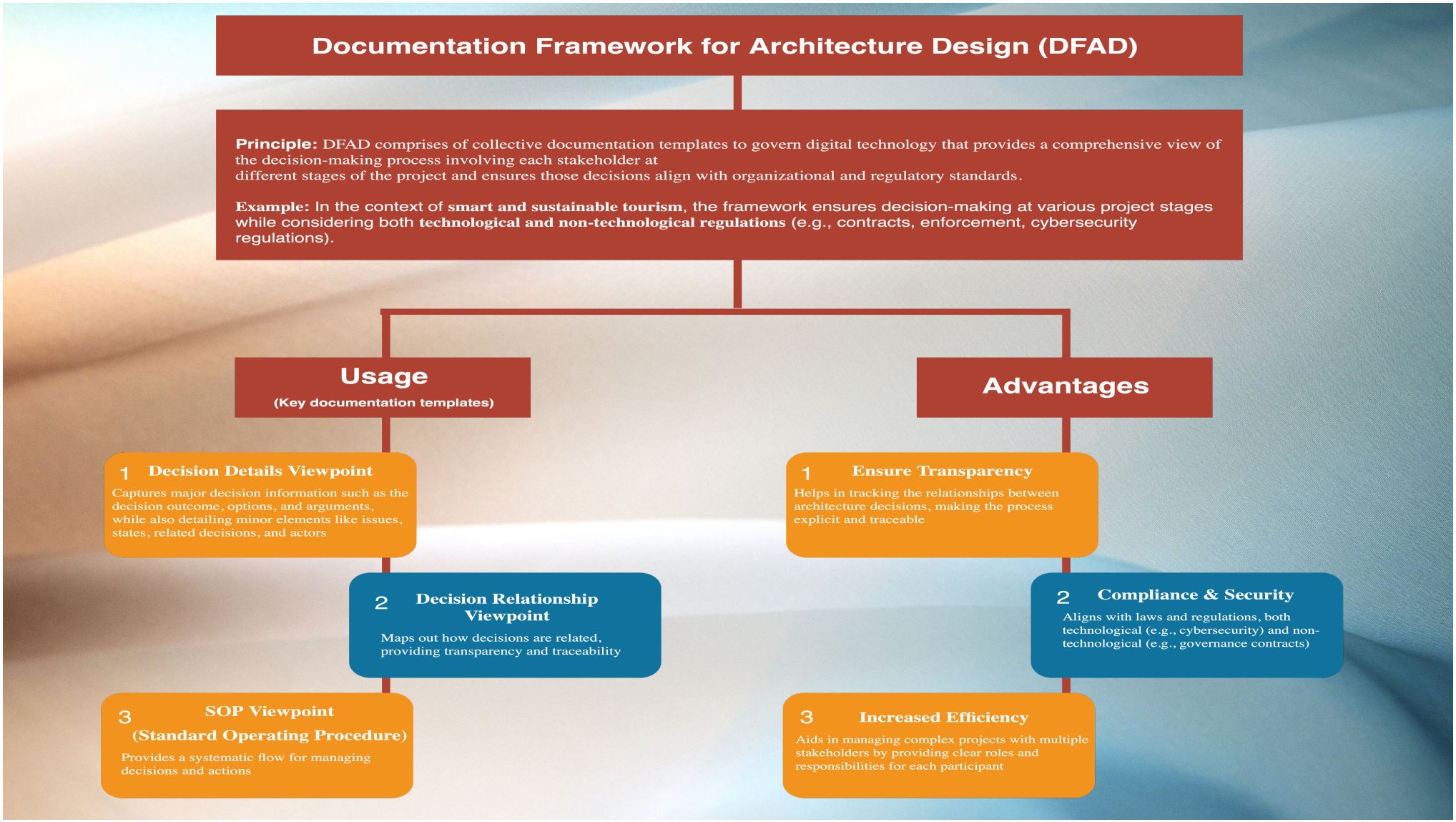
Ethics

- In the context of big data utilization, two notable ethical issues arise: privacy and data security
- For instance, certain applications have led to privacy and data security violations by exposing users' identities and locations
- Gotterbarn, 20002 highlights - a set of guidelines and regulations must be enabled by software specialists to prioritise the well-being, justice, and safety of users and societies

Framework

- A structured approach with policies, standards, tools, and processes
- Includes data scope determination, storage, security, and quality monitoring




Documentation Framework for Architecture Design (DFAD)
Principle: DFAD comprises of collective documentation templates to govern digital technology that provides a comprehensive view of the decision-making process involving each stakeholder at different stages of the project and ensures those decisions align with organizational and regulatory standards.
Example: In the context of smart and sustainable tourism, the framework ensures decision-making at various project stages while considering both technological and non-technological regulations (e.g., contracts, enforcement, cybersecurity regulations).
Usage
(Key documentation templates)
1 Decision Details Viewpoint
Captures major decision information such as the decision outcome, options, and arguments, while also detailing minor elements like issues, states, related decisions, and actors
2 Decision Relationship Viewpoint
Maps out how decisions are related, providing transparency and traceability
3 SOP Viewpoint (Standard Operating Procedure)
Provides a systematic flow for managing decisions and actions
Advantages
1 Ensure Transparency
Helps in tracking the relationships between architecture decisions, making the process explicit and traceable
2 Compliance & Security
Aligns with laws and regulations, both technological (e.g., cybersecurity) and non-technological (e.g., governance contracts)
3 Increased Efficiency
Aids in managing complex projects with multiple stakeholders by providing clear roles and responsibilities for each participant

DFAD Framework on DT technology for Indonesia Smart & Sustainable tourism implementation

Stakeholder	Role
Statistics Indonesia (BPS)	Data collection, script development, quality control
Mobile Network Operators (MNOs)	Data provider, ensuring security and compliance
Ministry of Tourism	Data user, policy recommendations
Legal Experts	Ensuring compliance with Indonesian privacy laws
Business Analysts	Bridging technical and business needs
Auditors & Regulators	Monitoring compliance and security

Framework

DFAD was applied to ensure accountability, transparency, and regulatory- compliance in decision making

- **Decision Details Viewpoint**
- **Relationship Viewpoint**
- **Standard Operating Procedures (SOP)**

Stakeholders & their roles

Case study involved **various actors**, each playing a role in governance and execution

Inferences

- **Enhanced Decision-Making**: DFAD provided **structured, transparent, and well-documented** decision processes.
- **Improved Big Data Governance**: Ensured **compliance** with **laws, security standards, and ethical guidelines**.
- **Stronger Stakeholder Coordination**: Defined clear roles and communication flows.
- **Better Data Accuracy & Reliability**: Improved MPD script led to **higher-quality tourism statistics**.

Stage 1

Data Source

- **Primary**: Mobile Positioning Data (MPD)
- **Secondary**: Google Mobility and User-Generated Data (UGC)

Stage 2

Stage 3

Key Decisions

Case study focused on **three main decisions**

Stage 4

Stage 5

Regulatory and Ethical considerations

Case study ensured compliance with:

- **Indonesian Law of Statistics (Act 16/1997)**
- **Personal Data Protection (Act 22/2022)**
- **Regulations on Electronic System Operators (2020)**
- **UN Fundamental Principles of Official Statistics**
- **Ethical AI & Data Governance Standards**

Privacy, security, and data **ethics** were major considerations, particularly due to **MPD's sensitivity**.

Stage 6

Stage 7

Challenges

- **Legal Complexity**: Navigating Indonesian data laws was challenging.
- **Stakeholder Buy-In**: Some actors were resistant to new documentation processes.
- **Technical Issues**: Ensuring MPD script improvements met high-quality standards.

Key Decisions Template

Decision 1

Use of MPD for Smart and Sustainable Tourism

- **Issue:** Determining if MPD could continue to be used as a data source for tourism statistics and DT technology
- **Decision:** Continue using MPD due to its historical success (since 2016) in monitoring mobility and tourism
- **Alternative:** Use other sources such as Google Mobility or UGC
- **Challenges:** Data availability, privacy concerns, administrative approvals

Decision 2

Improving MPD Data Processing & Script

- **Issue**: Errors in the existing MPD data script needed correction for better accuracy and reliability
- **Decision**: Improve the script and replace the old one
- **Alternative**: Continue using the old script temporarily
- **Challenges**: Ensuring script validation while maintaining data quality

Decision 3

Testing the Script in a Secure Environment (Sandbox)

- **Issue**: MPD cannot be processed outside the Mobile Network Operator (MNO) environment due to privacy regulations
- **Decision**: Test the new script in the MNO sandbox
- **Alternative**: Test Google Mobility or UGC data in Statistics Indonesia's data lake
- **Challenges**: Ensuring privacy, security, and compliance

Thanks much

Open to Questions...

We Are Group-2:
Surya Kurapati
Saranya Munikannaiah
Vishal Vishvanath Sawant Dessai
Rajesh Morthad