VTT Technical Research Centre of Finland

Strategies for data supply in high-granularity data trade in smart cities

Palviainen, Marko; Kotovirta, Ville

Link to publication

Strategies for data supply in high-granularity data trade in smart cities

Marko Palviainen[1] · Ville Kotovirta[1]

Abstract

The smart city infrastructures, such as digital platforms, edge computing, and fast 5G/6G networks, bring new possibilities to use near-real-time sensor data in digital twins, AR applications, and Machine-to-Machine applications. In addition, AI offers new capabilities for data analytics, data adaptation, event/anomaly detection, and prediction. However, novel data supply and use strategies are needed when going toward higher-granularity data trade, in which a high volume of short-term data products is traded automatically in dynamic environments. This paper presents offering-driven data supply (ODS), demand-driven data supply (DDS), event and offering-driven data supply (EODS), and event and demand-driven data supply (EDDS) strategies for high-granularity data trade. Computer simulation was used as a method to evaluate the use of these strategies in supply of air quality data for four user groups with different requirements for the data quality, freshness, and price. The simulation results were stored as CSV files and analyzed and visualized in Excel. The simulation results and SWOT-analysis of the suggested strategies show that the choice between the strategies is case-specific. DDS increased efficiency in data supply in the simulated scenarios. There was higher profit and revenues and lower costs in DDS than in ODS. However, there are use cases that require the use of ODS, as DDS does not offer ready prepared data for instant use of data. EDDS increased efficiency in data supply in the simulated scenarios. The costs were lower in EODS, but EDDS produced clearly higher revenues and profits.

Keywords Offering-driven data supply (ODS) · Demand-driven data supply (DDS) · Event and offering-driven data supply (EODS) · Event and demand-driven data supply (EDDS) · Data supply simulation · Air quality data

1 Introduction

There is a strong need for smart, sustainable, and energy-efficient cities. Achieving the United Nations' Sustainable cities and communities Development Goal (SDG 11) (UN 2022) requires solutions to enhance the decision-making and situational awareness in cities. In practice, this requires the development of systems-of-systems (Maier 1998) or cyber-physical systems (Robbins and Tanik 2014) to leverage the Data and Processing Capabilities (DPCs) in different kinds of systems in cities. This is challenging, as there are multiple stakeholders of DPCs in cities (Palviainen and Suksi 2023). Therefore, there is a clear need for data markets that

motivate data productization and value creation and enable efficient use of all kinds of data in a smart city for sustainable decision-making.

Some data have obvious value and established uses. For example, weather data are produced at the city level and sold to many applications via multiple distribution channels, including media. However, in a smart city, a lot of data in finer grain exist without established data markets, and data are not productized for other than the primary use. For example, air quality might be measured locally by smart lighting poles, building automation systems, vehicle air conditioning systems, and observed by individuals moving in the city. The data are scattered spatially and temporally in different formats and are not easily available for routing services based on air quality. To utilize these kinds of local, short-term, and moving data sources in novel services and create new value, we need to go toward higher-granularity data trade that includes data from multiple spatial and temporal scales.

✉ Marko Palviainen
marko.palviainen@vtt.fi

Ville Kotovirta
ville.kotovirta@vtt.fi

1 VTT Technical Research Centre of Finland, P.O. Box 1000, N 02044 Espoo, Finland

Creating value from data requires fair, secure, and efficient processes for data trade (Liang et al. 2018). There should be low market entry and exit costs and low transaction costs in data exchange (Duch-brown et al. 2017) and enough actors so that network effects (Fruhwirth et al. 2020) are achieved. In addition, there should be mechanisms to sell dynamic and frequently updating data for arbitrary ad-hoc queries (Liu and Hacigümüş 2014). The technological development set new requirements for data trade. The role of machines is rising in data trade, and it can be envisioned that systems such as autonomous vehicles and robots will use or supply sellable data, which refers here to data that are productized for transactions in data marketplaces. SDGs (UN 2022) should be achieved in data supply and use, too. The use of network and computing capabilities should be optimized to increase profitability and energy efficiency in the sellable data supply, delivery, and use. There should be situational and context awareness in data supply and data use, as the data demand depends often on the situation and context. For example, in a traffic jam, there is more demand for near-real-time traffic data that assist in navigation or in searching for a parking lot. The robotized data trade increases short-term supply and use of data in situations where a data source is available for only a relatively short time, e.g., when a car making air quality observations is driving through a city, and the data source can be used only for a limited time. Managing this dynamism requires increasing the granularity as well as situational and context awareness in data trade and novel strategies to optimize the sellable data supply and use.

From the literature, there can be found numerous studies that have explored increasing of the *efficiency*, *dynamism*, and *flexibility* in data trade. The efficiency is attempted to be improved by focusing on data markets' organization, trust, interoperability, and sellable data quality in data trade. The prior studies have proposed approaches for centralized and decentralized data markets (Ramachandran et al. 2018; Alsvåg et al. 2022; Anthony 2023) and for local, federated, and domain-specific data markets (Yerabolu et al. 2019; Fernandez et al. 2020; Gröger 2021; Abbas et al. 2022). There are efforts to build an infrastructure, rules, principles, and standards for achieving trust, interoperability, and data sovereignty in data sharing (Eggers et al. 2020). The Identity and Access Management (IAM) systems assist guaranteeing of the identity of data seller and data buyer (Mangiuc 2012; Ramachandran et al. 2018) and the blockchain technologies and smart contracts enable completing of transactions in an inexpensive and fast manner (Zheng et al. 2020). The open, standardized, and interoperable APIs are developed to assist collecting and processing of data from different sources in smart cities (Robin and Botts 2007; Park 2017; McGrath et al. 2019). In addition, there are approaches that enable automatic metadata generation for sellable data products

(Sharma et al. 2020) and tools that assist leveraging of big data in the smart city context (Khan and Kiani 2012; Badii et al. 2017; Anthony Jnr et al. 2020). The quality control, curation, and recommendation systems can be provided for sellable data to assist finding of good quality, reliable sellable data products in a marketplace (Ramachandran et al. 2018; Sharma et al. 2020). In addition, buyers and sellers can rate each other in a marketplace that assists the buyers to assess the quality of the data product provided by a given seller (Ramachandran et al. 2018). There can be dynamism in selling, pricing, processing, delivering, and using of data. There are dynamic pricing schemes and auction processes for data trade (Liang et al. 2018), mechanisms to share network capacity based on data demand (Lorenzo and Gonzalez-Castano 2016), and approaches to price and sell data that is composed for a data user's requirements (Liu and Hacigümüş 2014; Duch-brown et al. 2017; Fernandez et al. 2020). In addition, flexible terms of use, pricing models, and payment options are used to increase flexibility in data trade (Liang et al. 2018; Sharma et al. 2020).

However, although the above-mentioned studies have found ways to enhance efficiency, dynamism, and flexibility in data trade, there is still a need to increase granularity as well as market, situational, and context awareness in sellable data supply and use. This work extends the prior studies by introducing novel elements for data trade optimization. First, the Market, Situation, and Context Update (MSC)-messages are suggested to deliver fresh information about the market, situation, and context changes to enable continuous data supply and data use optimization in data trade. Second, ODS, EODS, DDS, and EDDS strategies are proposed to assist optimization of sellable data supply. On this basis, this study sets the following research questions: (a) How to increase granularity as well as market, situation, and context awareness in sellable data supply and use? (b) What kind of benefits and weaknesses relate to the suggested ODS, DDS, EODS, and EDDS strategies?

This paper makes the following contributions. First, a novel model is presented for high-granularity data trade to increase granularity as well as market, situational, and context awareness in data supply and use. Here, *market awareness* refers to mechanisms that actively optimize data supply for demand and data use for data supply. *Situational awareness* refers to sellable data products that adapt to changes in the physical environment, and *context awareness* means data products that adapt to changes in the processing environment, like edge processing capabilities, communication bandwidth, and electricity price. Second, offering-driven data supply (ODS), demand-driven data supply (DDS), event and offering-driven data supply (EODS), and event and demand-driven data supply (EDDS) strategies are presented for high-granularity data trade. Third, simulation

results and SWOT-analysis of the strategies are given, and the requirements that these strategies set for data products and marketplaces are discussed.

This paper is organized as follows: Sect. 2 provides literature review on related work. The characteristics of high-granularity data trade and the ODS, DDS, EODS, and EDDS strategies are then discussed in Sect. 3. The research methodology is discussed in Sect. 4 before the findings in Sect. 5 that discusses the simulation results. The discussion and implications are discussed in Sect. 6 and finally conclusions are given in Sect. 7.

2 Literature review

Data marketplaces can provide metadata descriptions for sellable data products and user interfaces and APIs for data trade that enable connection of organizations, data streams, hardware, and software into opaque and transacting collections (Taylor et al. 2022). Markets aggregate and express the knowledge of market actors (Hayek 1944) and use the pricing mechanism to disseminate information, channel dispersed resources, and fulfill people's localized needs and make the centralized economic planning unnecessary (Taylor et al. 2022). The following subsections focus on the drivers of smart city data economy, on data trading and sharing in smart cities, and on the previous efforts to optimize efficiency, dynamism, and flexibility in data trade. The final subsection presents a summary for the discussed data trade optimization approaches.

2.1 Drivers for smart city data economy

2.1.1 Urbanization

Since 1950, the number of large cities has increased very rapidly, and close to 400 cities now exceed a population of one million, and sprawling metropolitan areas have formed even larger agglomerations, and some urban regions with populations in the tens of millions have emerged (Berry et al. 2008).

2.1.2 Near real-time data

Delays in the processes that extract, integrate, and deliver data for users affect the data freshness that is identified to be one of the most significant data quality attributes (Bouzeghoub and Peralta 2004). The fast and reliable (e.g., 5G/6G) networks offer high uplink and downlink capacities that make the fluent delivery of high-resolution sensor data (e.g., image or video data) possible. This increases the data freshness, quality, and volume and makes the data more valuable for near real-time applications.

Data freshness can vary in autonomous data sources and integrating these data may lead to semantic problems in distributed systems (Bouzeghoub and Peralta 2004). The measuring of the different dimensions of data freshness can be based on the *currency metric* measuring the time elapsed since the source data changed without being reflected in the materialized view, on the *obsolescence metric* measuring the number of updates to a source since the data extraction time, and on the *freshness-rate metric* measuring the percentage of elements (tuples or attributes) that are up-to-date in data (Bouzeghoub and Peralta 2004). In addition, the *timeliness metric* is used to measure the extent to which the age of the data is appropriate for the task on hand (Wang and Strong 1996).

2.1.3 Digital systems

The following gives examples of systems that could supply or use sellable data in cities:

- *Smart city infrastructures* There can be smart urban furniture (Tokuda 2004; Krejcar et al. 2019), smart poles, and sensors such as cameras, lidars, and air quality sensors to produce street-level sensor data (Di Vito et al. 2020). These systems can be connected to marketplaces to offer data, edge computing capabilities, and event and anomaly detection capabilities (e.g., (Xu et al. 2020) for different actors in cities.

- *Edge computing* enables the processing of data at the edge of the network, at the proximity of data sources and end users, and provides the means to enhance response times, energy efficiency, and security in data processing (Shi et al. 2016; Hassan et al. 2019). Edge computing can be used together with cloud services to process downstream data on behalf of cloud services and upstream data on behalf of the Internet of Things (IoT) services (Shi et al. 2016).

- *Connected and automated vehicles (CAVs)* have a great potential to improve road safety, quality-of-life, and the efficiency of transportation systems (Elliott et al. 2019). CAVs can act as data suppliers, as they can produce sensor data of physical environment and benefit from data (e.g., RGB-D data and point cloud data) that is produced in the smart city infrastructure or in other vehicles in the city.

- *Robots* will facilitate everyday life, manipulate the environment, supply services, inspect and maintain infrastructure, build new structures in a city, or interact with citizens (Tiddi et al. 2020). For example, there can be assistive robots in homes, robots that participate in the business operations of cities, and robots, such as flying drones, to gather sensor data in cities (Studley and Little 2021).

- *Smart buildings* use IoT sensors, actuators, and solutions for building management that can assist in heating, cooling, ventilation, lighting, and water management and help increasing the safety ratio and energy efficiency in buildings (Verma et al. 2019).

2.1.4 Near real-time applications

Situational awareness and prompt decision-making are relevant in complex urban environments. Near-real-time applications require data products that can be purchased easily and quickly, processed without delays, and delivered fast to end users—all automatically without manual operations.

2.2 Data trading and sharing in smart cities

Data are considered as one of the most valuable assets that enables creation of data-driven services ranging from transportation and safety to health and sanitation to continually improve the lives of citizens (Docherty et al. 2018; Ramachandran et al. 2018). Data are shared as freely available open data in cities. However, although there is much interest around open data and data brokerages, only a small amount of available data from cities and public authorities are open data and there is a lack of clear use cases to leverage on data and develop innovative services for the benefit of a community (Alvsvåg et al. 2022).

The European Union (EU) has started to advocate for the adoption of open, standardized APIs with interoperable, coherent protocols and formats for collecting and processing of data from different sources in smart cities (McGrath et al. 2019). One of the goals is to support interoperability of datasets for a thriving smart city data-driven economy that creates digital innovations and services for sustainable development. The data exchange is often based on Representational State Transfer (REST) APIs and MQTT (Message Queuing Telemetry Transport). The REST APIs allow the dynamic pull of data from databases in response to end user's requests or inputs, instead of pushing same static information to every user. These APIs are based on the REST protocol (Fielding 2000) and a stateless procedure which suggest that each call comprises of information essential for execution and thus does not require status information from prior calls (Karnouskos et al. 2012). MQTT, in turn, is an extremely lightweight and simple messaging protocol that offers publish/subscribe mechanisms for data exchange (Anthony Jnr et al. 2020). MQTT is designed for low-bandwidth constrained devices that employs unreliable networks and, for example, can be used for connecting energy metering devices in energy district due to its small headers and minimum overhead (Anthony Jnr et al. 2020). Moreover, MQTT can also

be deployed over Secure Socket Layer (SSL) to implement security (Patti and Acquaviva 2016).

Data sharing and trade can relate to big data requiring tools to manage volume, velocity, veracity, and variety of the data (Khan and Kiani 2012). For example, the physical devices in energy districts will produce terabytes of data (Li et al. 2017) that can enable forecasting of energy markets, detecting usage anomalies for early warning, and offering recommendations and guidance for prosumers and energy companies (Badii et al. 2017). However, implementation of these services requires solutions such as a layered architecture and APIs presented in Anthony Jnr et al. (2020) to provide an access to the real time and historical energy data.

Data are scattered across different platforms in cities, and it is hard to find the data and there can be different access mechanisms for the data (Alvsvåg et al. 2022). One solution to solve this problem is to have a service-oriented and data-driven IoT software architecture for smart cities (Simmhan et al. 2018). Data marketplaces assist in data discovery and reuse in cities, offer incentives for data sharing, and enable data sellers to create value and perhaps earn money from their data (Alvsvåg et al. 2022). Data trade can be based on centralized or decentralized data marketplaces in smart cities. For example, a smart city data marketplace can be a digital platform enabling easy selling, buying, and sharing of data that can be Internet of Things (IoT) sensor, citizen, and business data from the smart city (Alvsvåg et al. 2022). Operators of centralized data marketplaces have significant monopoly market power over how data products are presented to the data consumers (Ramachandran et al. 2018). Decentralized data marketplaces minimize this risk of a biased marketplace operator that favors certain data suppliers by offering better visibility and rankings for their sellable data products. In addition, decentralized trust mechanisms such as storing and retrieving metadata and ratings from a Blockchain can potentially improve the transparency and trustworthiness of ratings (Ramachandran et al. 2018).

2.3 Methods for data trade optimization

There are developed solutions to increase efficiency, dynamism, and flexibility in data trade. These elements are discussed in the following paragraphs.

2.3.1 Efficiency in data trade

The integration of blockchain technology with smart contracts provides a way for finishing transactions in an inexpensive and fast manner without the need for a trusted third party (Zheng et al. 2020). The smart contracts consist of contract clauses that are executed when predefined conditions are met (Zheng et al. 2020). The smart contracts are stored to distributed blockchains assuring the distributed

trust, as it is nearly impossible to tamper with transactions stored in blockchains, and all the historical transactions are auditable and traceable (Zheng et al. 2020).

The better organization of data markets can improve the extraction of value from data. For example, domain-specific markets (markets for finance, for health, for agriculture) may be more efficient than more general ones in detecting highly valuable datasets (Fernandez et al. 2020). The GAIA-X and i3-Market initiatives aim at creating a single European Data Market by 2030. The objective is to amplify network effects in data markets by enabling data suppliers to join a meta-platform to make their data visible for the data users in the participating marketplaces (Abbas et al. 2022). Data sovereignty must be ensured so that data owners can exclusively decide which data they want to share with whom, for how long, under what conditions of use, and at what price (Bornholdt 2021). The GAIA-X project focusses on this and builds an infrastructure, rules, principles, and standards for achieving trust, interoperability, and data sovereignty in data sharing (Eggers et al. 2020). The International Data Spaces (IDS) standard is integral to GAIA-X enabling use of the IDS connectors that put the shared data inside a trusted, certified data space, ensuring that it is used only as agreed upon per the terms set by the parties involved (IDSA 2019).

The data marketplaces can focus on local users and use cases (e.g., in geofenced areas or buildings) requiring high trust, security, and speed in data exchange. For example, (Anthony 2023) presents recommendations for using decentralized data marketplaces to contribute to sustainable development in cities. The sellable data products can be based on Edge servers' Data and Processing Capabilities (EDPCs) that keep the data at the local level and do not transfer the "raw" or unprocessed data to the cloud (Palviainen and Suksi 2023). There are developed enterprise data marketplaces that contain a metadata-based inventory for EDPCs in edge data lakes to enable the realization of applications based on local data (Gröger 2021). However, these marketplaces focus typically more on matching data supply and demand within the enterprise than providing data for external users (Yerabolu et al. 2019).

The national and EU-level data-related regulation is evolving, and there is a vast amount of data-type-specific, e.g., personal data under GDPR (EU 2016) and sector-specific regulations and the upcoming Data Governance Act (EU 2022), Data Act (European Commission 2022), and AI Act (European Commission 2021) that must be considered in data exchange and trade (Palviainen and Suksi 2023). Data productization requires legal and contractual frameworks to decrease legal uncertainty regarding trading data (Duch-brown et al. 2017; Spiekermann 2019; Bornholdt 2021).

The metadata describing the sellable data products in marketplaces can be organized hierarchically, tagged in some way, or left unstructured (Ramachandran et al. 2018) and standards such as SensorML (Robin and Botts 2007) or OCF (Park 2017) data models offer standardized ways to organize the sensor information. Automatic metadata generation (see, Sharma et al. 2020) increases efficiency in data productization, as it reduces the need for manual metadata production, decreases errors in metadata, and offers a way to produce more metadata to assist in selection of data products.

The quality control, curation, and recommendation systems can be provided for sellable data to assist finding of good quality, reliable sellable data products in a marketplace (Ramachandran et al. 2018; Sharma et al. 2020). For example, Sharma et al. (2020) discuss a service that enables a data buyer to ensure in the purchase phase that the data set fulfills the desired quality requirements. In addition, buyers and sellers can rate each other in a marketplace that assist the buyers to assess the quality of the data product provided by a given seller (Ramachandran et al. 2018). In addition, a simulation approach (Palviainen and Kotovirta 2021) is developed to assist optimization of vehicle-based supply of data for quality and cost-oriented data users.

2.3.2 Dynamism in data trade

Digital technologies and algorithms offer ways to decline the cost of collecting, distributing, searching, and using data coming from various sources (Duch-brown et al. 2017). Data trading schemes assist in the data pricing, and there can be an auction process to allocate commodities and establish corresponding prices through data buyers' and data sellers' bidding process (Liang et al. 2018). There are marketplaces that allow the dynamic use of data processing capabilities. For example, a DeepMarket marketplace enables users to lend or borrow edge computing resources for the distributed execution of ML programs (Yerabolu et al. 2019). There can be dynamism in data delivery, too. For example, (Lorenzo and Gonzalez-Castano 2016) discuss an approach to incentivize users to share their connectivity in a user-provided network and obtain a profit by selling and buying leftover data capacities (caps) from each other.

2.3.3 Flexibility in data trade

There can be flexibility in data offerings, payment methods, and pricing models. For example, there is the concept of using arbiters in the combination of individual datasets to add value in different mashups to satisfy a varied set of buyers' needs (Fernandez et al. 2020). Data suppliers can offer different pricing models, such as free data (e.g., sample data), usage-based pricing (e.g., data stream usage and service time-based pricing), or package pricing models (e.g., a data package plan with a fixed price) for sellable data (Liang

et al. 2018) that allows data users to select the most appropriate pricing model for data use. The data marketplaces can offer different payment options and, for example, (Sharma et al. 2020) discuss a data marketplace enabling paying with cryptocurrencies, PayPal, and traditional banking. For example, the use of cryptocurrencies enables seamless micro-payments for data making it possible to perform more flexible and shorter-term interactions (Ramachandran et al. 2018).

2.4 Related works

Table 1 presents the prior studies that have developed solutions to increase efficiency, dynamism, and flexibility in data trade. This work extends the previous studies by introducing novel elements for data trade. First, the Market, Situation, and Context Update (MSC)-messages are proposed to deliver fresh information about the market, situation, and context changes to enable continuous data supply and data use optimization in data trade. Second, the ODS, EODS, DDS, and EDDS strategies are proposed to assist optimization of the sellable data supply.

3 Four strategies to supply data in high-granularity data trade

Technological development, such as robotized data trade, sets new requirements for data trade. For example, the use of network and computing capabilities should be optimized so that sustainable development, profitability, and energy efficiency goals are achieved in the sellable data supply, delivery, and use. The rest of this paper focuses on high-granularity data trade that aims at achieving these goals by increasing granularity as well as situational and context awareness in sellable data supply and use optimization.

Although there are previous efforts to increase the efficiency, dynamism, and flexibility in data trade (in Table 1), the existing approaches do not focus on increasing market, situational, and context awareness in sellable data supply and use optimization. This section presents a novel model for high-granularity data trade to increase granularity as well as market, situation, and context awareness in sellable data supply and use. By achieving these goals, it is possible to further improve (see Fig. 1):

(1) *Efficiency in data trade* The Market, Situation, and Context Update (MSC)-messages increase profitability and energy efficiency in data exchange and trade, as they deliver fresh information about the market, situation, and context changes to enable continuous data supply and data use optimization.

(2) *Dynamism in data trade* The short-term data supply and short-term data use increase dynamism in data trade. This requires that it is fluent to join to marketplaces. There should be low latencies in data discovery, exchange, and trade and ways to sell and buy data without human intervention.

(3) *Flexibility in data trade* There are data supply and data use strategies for different purposes and use cases to increase flexibility in data trade. Multiple pricing models are offered for data and the changes in data freshness, quality, and volume can affect pricing and selection of data products.

Achieving these goals require building of an adaptive system for sellable data supply and use. This kind of system can be based on a loop that delivers MSC-messages for data use and data supply optimization (see the upper part in Fig. 2). This work focuses on the data supply optimization part and considers *data offerings* and *data demand* as main drivers on data markets. There should be data supply strategies for these drivers. First, there should be strategies that assist supply of data offerings for the expected demand, context, and situation. Second, it should be possible to supply adapted data for data users' requirements, context, and situation. We used these requirements and the optimization loop in Fig. 2 as a starting point and developed strategies for the following:

(1) *Offering-driven data supply (ODS)* that delivers fully prepared data for data buyers. A data supplier prepares the data and publishes it as a sellable data product in the marketplace and trusts that there is demand for it. There are (at least) two options for ODS. A data supplier can use ODS_p (p for periodic) and perform periodic updates for data offering. Or a data supplier can use ODS_{max} and attempt to maximize its market share in the marketplace. In ODS_{max}, the data supplier can estimate the market share by first computing the number of active data users for which the supplier offers the most suitable data product in the marketplace and update the data offering if it increases that number.

(2) *Demand-driven data supply (DDS)* that delivers adapted data for data users' requirements: a data supplier prepares data supply capabilities, then publishes a *showcase data product* in the marketplace, and finally prepares the data after a data user has bought the data product and request the data for use. The data can be adapted and delivered at a desired resolution and quality level for decreasing the data volume, use of network connections, and data transmission latencies.

(3) *Event and offering-driven data supply (EODS)* that delivers situational and contextual data for detected events: a data supplier prepares the situational data and publishes it as a sellable data product in the marketplace and trusts that there is demand for the data.

Table 1 Approaches to increase efficiency, dynamism, and flexibility in data trade

	Goal 1: Efficiency in data trade	Goal 2: Dynamism in data trade	Goal 3: Flexibility in data trade
Existing approaches and methods	• Centralized or decentralized data marketplaces (Alvsvåg et al. 2022; Ramachandran et al. 2018; Anthony 2023)	• Dynamism in data offerings (Duch-brown et al. 2017; Fernandez et al. 2020)	• Flexible data offerings (Fernandez et al. 2020)
	• Local data markets (Anthony 2023; Palviainen and Suksi 2023; Gröger 2021; Yerabolu et al. 2019)	• Dynamism in data pricing (Liang et al. 2018)	• Flexible pricing models (Liang et al. 2018)
	• Federated data markets (Abbas et al. 2022; Eggers et al. 2020; IDSA 2019)	• Dynamism in data processing (Yerabolu et al. 2019)	• Flexibility in payment options (Sharma et al. 2020)
	• Domain-specific data markets (Fernandez et al. 2020; IDSA 2019)	• Dynamism in data delivery (Lorenzo and Gonzalez-Castano 2016)	
	• Identity and Access Management (IAM) systems (Mangiuc 2012; Ramachandran et al. 2018)		
	• Smart contracts and blockchains (Zheng et al. 2020)		
	• Regulation (EU 2016, 2022; European Commission 2022; European Commission 2021)		
	• Legal and contractual frameworks (Bornholdt 2021; Duch-brown et al. 2017; Spiekermann 2019)		
	• Standardized and interoperable APIs, protocols, and architectures for data sharing (McGrath et al. 2019; Fielding 2000; Anthony Jnr et al. 2020; Patti and Acquaviva 2016; IDSA 2019; Simmhan et al. 2018)		
	• Tools for big data (Khan and Kiani 2012)		
	• Automatic metadata generation and metadata standards (Sharma et al. 2020; Ramachandran et al. 2018; Robin and Botts 2007; Park 2017)		
	• Quality controls for sellable data (Sharma et al. 2020)		
	• Curation and recommendation for data and ratings for data sellers and data buyers (Ramachandran et al. 2018)		
	• Simulations (Palviainen and Kotovirta 2021)		
Methods for high-granularity data trade	• Market, Situation, and Context Update (MSC)-messages		
	• Situational and context awareness in data supply and data use		
	• Offering-driven supply of data		
	• Demand-driven supply of data		
	• Event and offering-driven supply of data		
	• Event and demand-driven supply of data		

The situational data can assist in decision-making and enable faster reaction to detected events and anomalies.

(4) *Event and demand-driven data supply (EDDS)* that delivers adapted situational and context data for detected events: a data supplier prepares data supply capabilities, then publishes a showcase data product in a marketplace, and finally prepares the data after a data user requests the data for use.

Following subsections analyze these four data supply strategies in more detail.
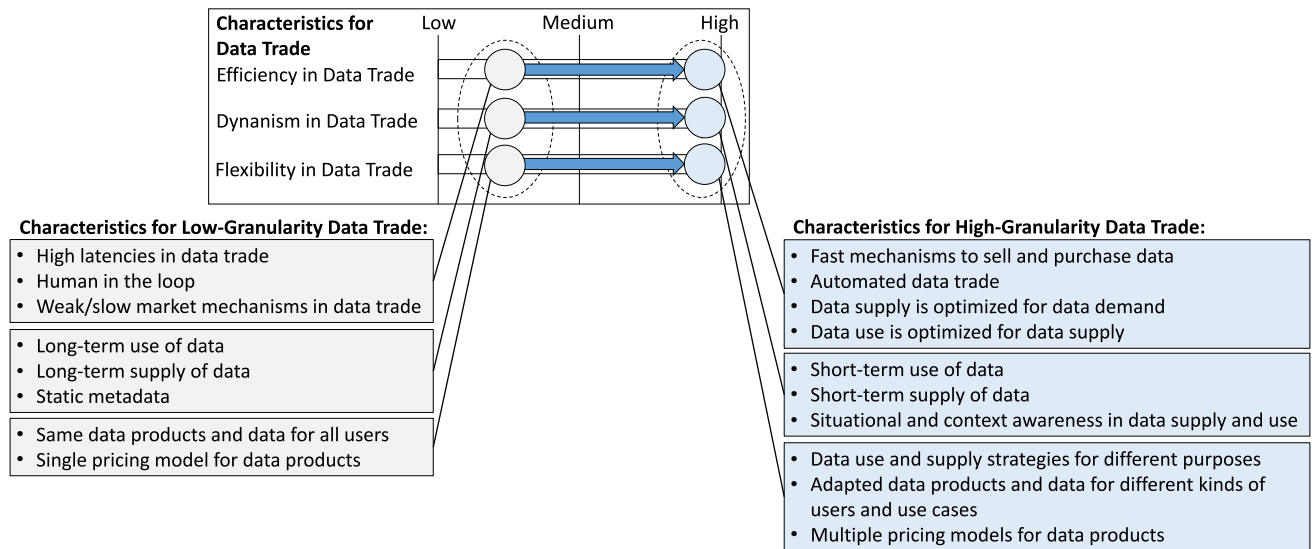
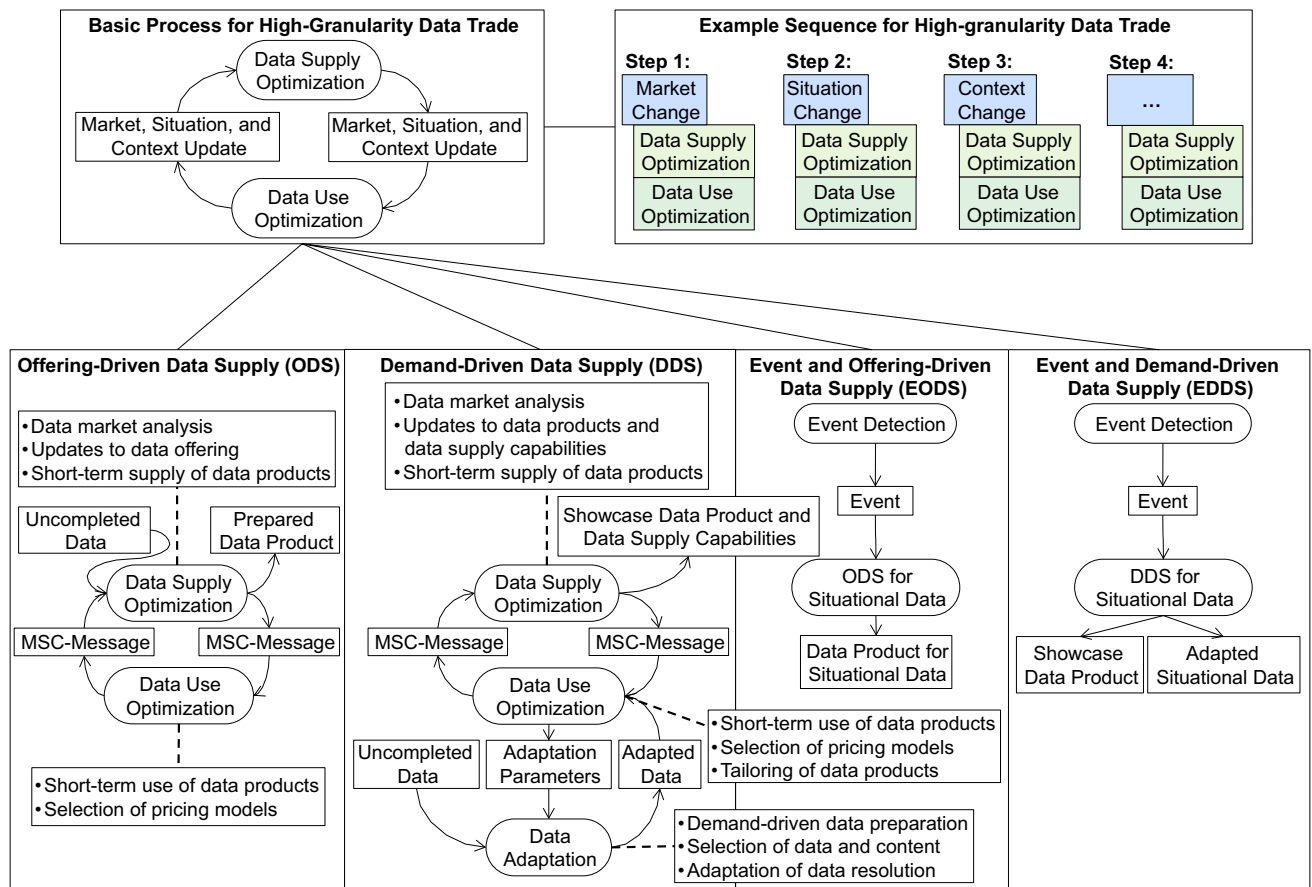**Fig. 1** The characteristics for high-granularity data trade

Fig. 2 A basic process, example, and four data supply strategies for high-granularity data trade

3.1 Efficiency in data supply

The proposed strategies increase efficiency in data supply by optimizing the data supply for the estimated demand (see Fig. 2). The MSC-messages deliver fresh information about data users and sellable data products in the marketplace. MSC-messages can contain data supply notifications (DSNs) and data use notifications (DUNs). DSNs offer information about data products' prices, volume, freshness, and quality. DUNs deliver information about data users' requirements for data and about new, interrupted, or canceled data product subscriptions. In addition, DUNs contain requests for new data products if there are gaps in the sellable data coverage and availability.

In demand estimation, there can be used prediction methods, information delivered in MSC-messages, and other market information, such as historical and statistical information about data users and data use. The demand estimation can analyze the data demand, data supply, and data user distribution in the marketplace, and identify the data users' requirements for data quality, price, and freshness.

The data supply is optimized in two phases in DDS and EDDS. The data supply capabilities and showcase data products are first prepared for the estimated demand. The data are prepared after there is a user for the data. ODS and EODS prepare the sellable data directly for the estimated demand. However, the data are ready to be used, but there is a risk that there are not enough users for the data if the demand is estimated incorrectly.

3.2 Dynamism in data supply

The proposed strategies offer ways for increasing dynamism in data supply. The short-term supply of data is based on MSC-messages that assists the data suppliers to analyze the markets and to adapt data supply when there are gaps in data availability (e.g., in high-quality data availability) or interrupt or stop the data supply if there are overlapping (e.g., cheaper) data products in the marketplace. Adapting to demand changes is easier in DDS than in ODS. Instead of preparing the data for changing demand, it is faster and cheaper to publish showcase data products to make the potential data offering visible in the marketplace and prepare the data after there are users for it.

3.3 Flexibility in data supply

The proposed strategies increase the flexibility in data supply. ODS and EODS offer data for the expected demand. DDS and EDDS can deliver adapted data for the data user's requirements. In addition, a data user can tailor the data product to contain the desired data elements, features, content, information, and functionalities, reducing data use cost in the purchase phase.

Alternative pricing models for the data products can be provided, such as API call-based pricing models for occasional use of data and subscription-based pricing models for continuous use of data. In EODS and EDDS, there can be also event-related pricing models for data. For example, the delay that it takes to offer data for an event may affect the pricing of data.

4 Methodology

Computer simulation was used as a methodology to evaluate the use of the suggested ODS, EODS, DDS, and EDDS strategies in the supply of air quality data for four user groups with different requirements for the data quality, freshness, and price. The simulation followed the basic structure of computer simulation that consist of (a) a *model* that defines the structure and behavior of the system that is the subject to the study, (b) a *scenario* that is a model of the exogenous stimuli applied to the system, (c) a *simulator* that refers to a simulation engine that is reusable for the simulation of many models and scenarios, and an (d) *instrumentation* that defines which variable(s) of interest to observe during the simulation execution, what computation(s) to apply at runtime to the data samples produced by these variables, and what is logged to result files (Dalle 2012).

A simulator prototype that is implemented in JavaScript was used in the simulation. The simulator prototype was used in a Microsoft Edge browser on a laptop computer with a 2.10 GHz Intel Core i7-12850HX processor and 64 GB of main memory. The simulation results were stored as CSV files and analyzed and visualized in Excel. The following subsections discusses the model and parameters that were used in the simulations.

4.1 Simulation model

The simulations evaluate the supply of air quality data in a city (see Fig. 3) that has a smart lighting pole infrastructure, data platform, marketplace, and 5G connections for data exchange and trade.

The Air Quality Index (AQI) (FMI 2023) used in Finland considers the concentrations of sulfur dioxide (SO_2), nitrogen dioxide (NO_2), respirable particles (PM_{10}), fine particles ($PM_{2.5}$), ozone (O_3) carbon monoxide (CO), and the total reduced sulfur compounds (TRS). The smart lighting poles are equipped with air quality sensors and provide a service that delivers notifications of AQI change events for data users.

Two kinds of data products are on sale in the marketplace:

Fig. 3 A use case for high-granularity data trade in a smart city
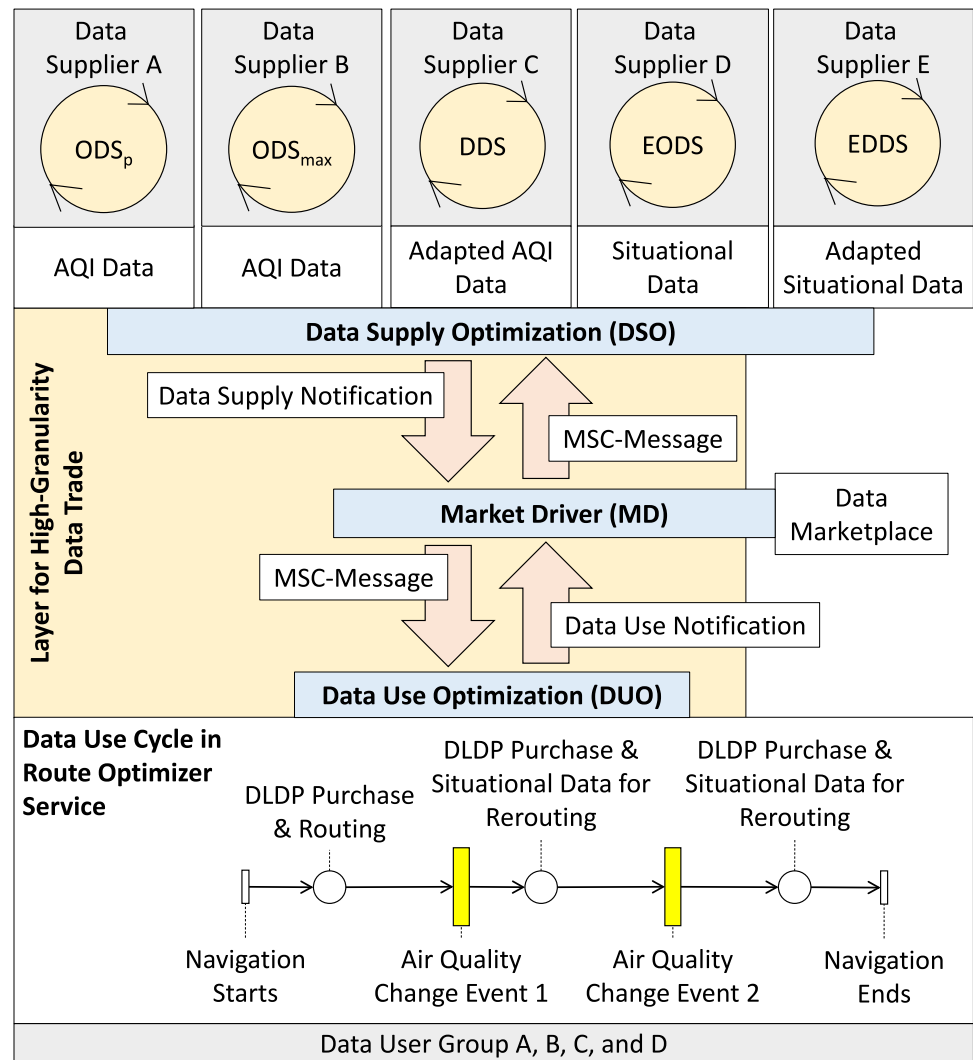
(a) *Ultra-Local Data Products (ULDPs)* that offer pole-specific air quality data for a single measurement point in the city. ULDP provides computed AQI for a single pole, measured values for SO_2, NO_2, PM_{10}, $PM_{2.5}$, O_3, CO, and TRS, and a classification for the air quality.

(b) *District-Level Data Products (DLDPs)* The data suppliers purchase ULDPs, use the ultra-local air quality data in computation of AQIs for the street sections in a specific city district, and finally prepare DLDPs for the marketplace to provide AQI data, adapted AQI data, and situational data or adapted situational data for data users. The data suppliers produce high-quality DLDPs to offer under 1-min-old AQI data and low-quality DLDPs to offer 1- to 3-min-old air quality data for a specific district in the city.

The data users are pedestrians and motorists that use a *Route Optimizer Service* for navigation in the city. The service computes route options and uses air quality data

(DLDPs) for selecting the route with the best air quality for the user. In a use cycle, DLDP is first purchased for a route selection (see Fig. 3). If AQI change events occur, the service purchases DLDP for situational data and calculates an updated route for the data user.

The high-granularity data trade features are implemented as a layer that consists of the market driver (MD), data use optimization (DUO), and data supply optimization (DSO) components (see Fig. 3). The DSO components drive data supply and DUO components drive data use. The MD component monitors the data trade in the marketplace, composes the DSNs and DUNs to MSC-messages, and finally delivers these messages for the DUO and DSO components.

4.2 Simulation parameters

Table 2 depicts the parameters used in the simulations.

Figure 4 is based on the traffic census numbers in Lilleberg and Hellman (2015) and shows how the daily vehicle

Table 2 Simulation parameters

	Data supplier A	Data supplier B	Data supplier C	Data supplier D	Data supplier E
Data supply strategy	ODS_p	ODS_{max}	DDS	EODS	EDDS
Sellable data	AQI Data for Routing		Adapted AQI Data for Routing	Situational Data for Rerouting	Adapted Situational Data for Rerouting
Pricing for high-quality DLDPs	€1/DLDP		€1.1/Data Request	€1/DLDP	€1.1/Data Request
Pricing for Low-quality DLDPs	€0.3/DLPD		€0.4/Data Request	€0.3 / DLPD	€0.4/Data Request
Input data cost	5 cents for high-quality DLDP and 2 cents for low-quality DLDP				
Data preparation cost	€0.01/DLDP		–	€0.01 / DLDP	–
Data adaptation cost	–		€0.01/Data Request	–	€0.01/Data Request
User Group A	100 data users that want to use high-quality DLDPs and the most up-to-date data				
User Group B	100 data users that want to use low-quality DLDPs and the most up-to-date data				
User Group C	100 data users that want to use high-quality DLDPs that offer the lowest price for data				
User Group D	100 data users that want to use low-quality DLDPs that offer the lowest price for data				
Data use cycle duration	10 min				
Simulation time	24 h				
Time step size	1-s step size				

Fig. 4 The diagram is based on the traffic census numbers in Lilleberg and Hellman (2015) and presents the percentage of daily traffic during different hours on working days in the city of Helsinki
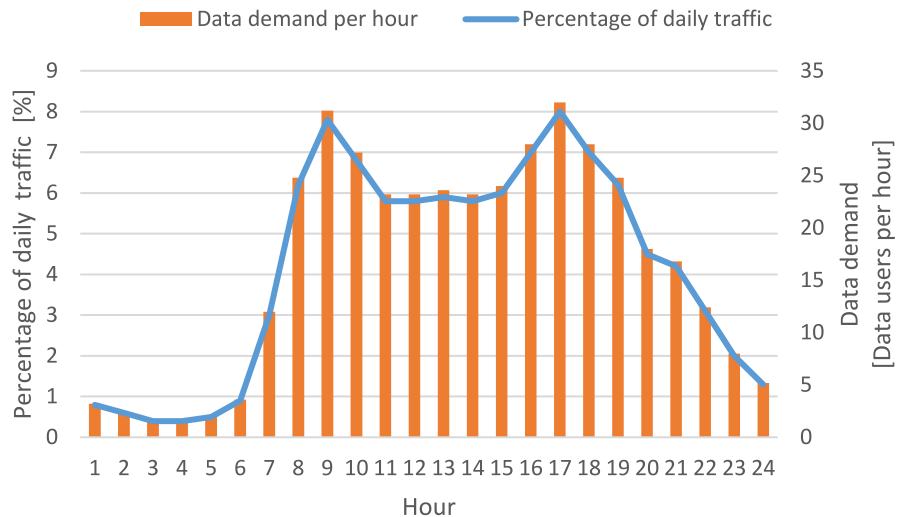
traffic is spread for different hours on working days in the city of Helsinki. These numbers were used as a starting point for the simulations that focus on four data user groups (see Table 2) that each consists of 100 users and have different requirements for the data quality, freshness, and price. The number of data users per hour (the demand for data) follows the hourly distribution of the vehicle traffic in Helsinki (see Fig. 4) so that in each simulation step there is same number of users from each four data user group. The data users move on a well-defined geographical region, and it is assumed that each DLDP offers the needed data for the route selection.

It is assumed that the time to use the *Route Optimizer Service* (the use cycle duration) is 10 min for each data user. The DUO component attempts to maximize the data quality and data freshness for groups A and C and minimize the data use cost for groups B and D by purchasing the cheapest DLDPs in a correct quality category for these users. The DUO component searches first DLDPs that offer AQI data with the sufficient quality for the computed route options. The DLDP that offers the newest or the cheapest data in the quality category is then purchased. The short-term use and replacement of data products, pricing model selection, data product tailoring, and data adaptation could be used in data use optimization, too. However, this paper focuses on data supply optimization, and for simplicity, the data product selection is the only data use optimization technique that is used in the simulations.

4.2.1 Costs

ULDPs offer input data for DLDP preparation. It is assumed that the input data cost is five cents for high-quality DLDP and 2 cents for low-quality DLDP. Data preparation cost includes data processing and storing cost and is one cent for each DLDP in simulations. The cost of adapting the data for a data user's request is one cent for a data request. Please note that the earnings, prices of data requests, and input data are not based on real pricing examples, as market information is not easily available. However, arbitrary heuristic values can be used, as the main point of the simulations is not to estimate absolute realistic profit values, but to compare the different strategies.

4.2.2 Pricing in ODS and EODS

The data marketplace can take a commission of sold data products. The revenues depend on the data quality, and it is assumed that a data supplier earns €1 from a high-quality DLDP and €0.3 from a low-quality DLDP (see, Table 2).

4.2.3 Pricing in DDS and EDDS

The data request price is €1.1 in high-quality DLDPs and €0.4 in low-quality DLDPs. In simulations, one data request in each transaction is made and, thus, a data supplier earns either €1.1 or €0.4 from each transaction in DDS and EDDS.

5 Findings

Simulations evaluate data suppliers' profits, revenues, and costs in two simulation scenarios. The first scenario compares offering and demand-driven strategies without event processing (ODS and DDS), while the second scenario considers the strategies that use events (EODS and EDDS). The time step affects the resolution of the simulation results. The time elapsed in a 24-h duration simulation by using a 1-s time step in scenarios 1 and 2 was 1 h 30 min and 54 s. The results from simulation scenarios 1 and 2 are discussed in the following subsections.

5.1 Results from simulation scenario 1

The simulation scenario 1 focuses on three data suppliers that use the following strategies in AQI data supply:

(1) *Data supplier A* relies on ODS_p and performs periodic updates for the data offering (DLDPs). The AQI data update time is in range of 30 s to 5 min in simulations.
(2) *Data supplier B* relies on ODS_{max} and attempts to maximize its market share in the marketplace.
(3) *Data supplier C* relies on DDS. If possible, the previously purchased input data are used in preparation of AQI data for new requests. The preparation of high-quality AQI data requires less than 1-min-old input data and the preparation of low-quality AQI data requires less than 3-min-old input data.

Scenario 1 consists of the sub-scenarios 1a, 1b, and 1c. In scenario 1a, there is only one data supplier and no competition in data supply (in Figs. 5 and 6). In scenario 1b, there are two rivals and, in scenario 1c, there are three rivals in data supply (in Fig. 6).

5.1.1 Scenario 1a

Figure 5 shows simulation results for data supplier A in scenario 1a. As can be seen, the longer AQI data update times decrease the cost but also revenues, as the AQI data available for data users are not fresh enough. The maximum profit of

Fig. 5 Profit, revenues, and cost for data supplier A in simulation scenario 1a
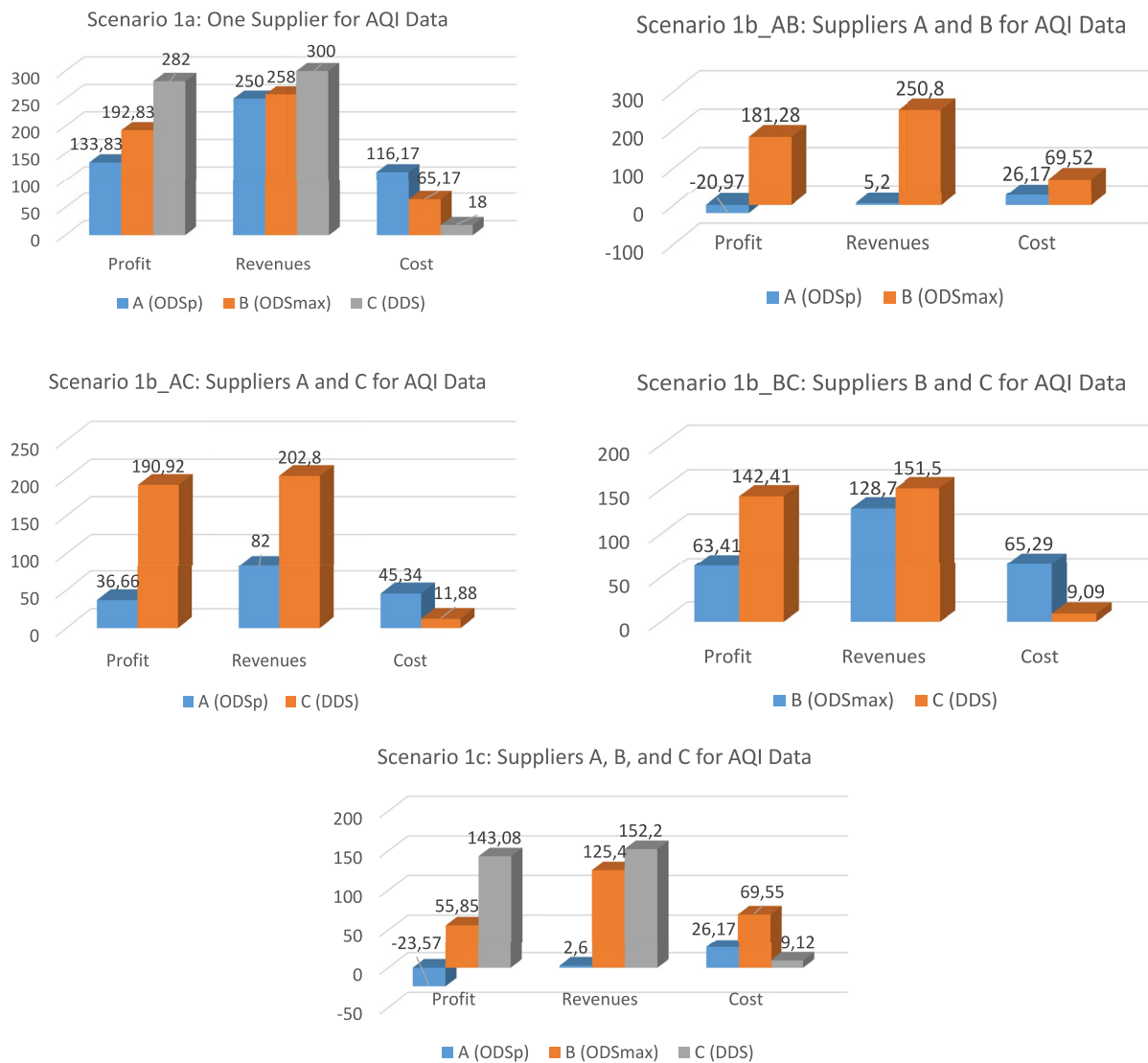

1a: Data Supplier A (ODS_p)
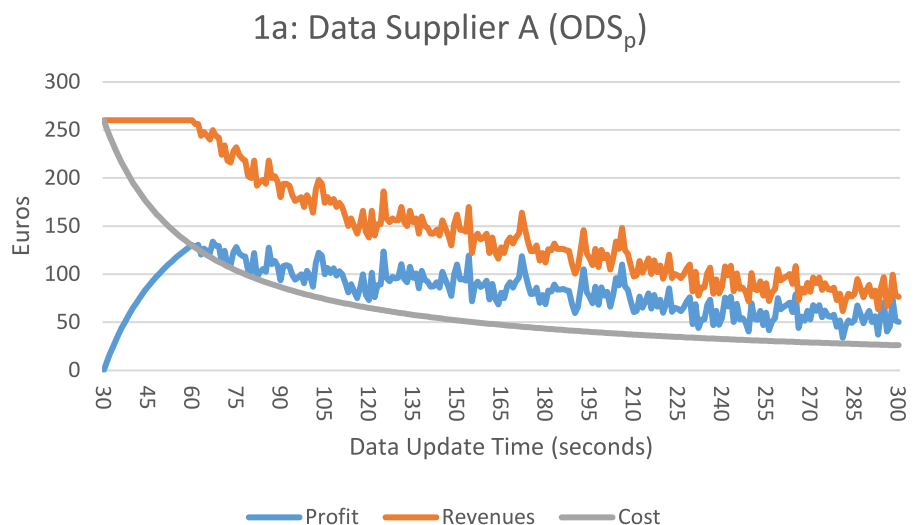
Fig. 6 Profit, revenues, and cost for data suppliers A (ODS$_p$), B (ODS$_{max}$), and C (DDS) in the simulation scenarios 1a, 1b, and 1c

€133.83 is achieved when the AQI update time is 1 min and 7 s. Data supplier B updates the AQI data for maximizing its market share and achieves almost 100% market share, as it is the only data supplier in scenario 1a. Data supplier C performs demand-driven supply of AQI data and saves in data preparation costs (see Fig. 6). The profit is €282 for data supplier C, and €192.83 for data supplier B in scenario 1a.

5.1.2 Scenario 1b_AB

Data suppliers A and B are competing in scenario 1b_AB. Supplier A achieved the highest profit (or the lowest loss) -€20.97 when the AQI data update time was 4 min and 58 s (see Fig. 6). In this case, the profit is €181.28 for data

supplier B. Supplier A has a lower cost, but supplier B has significantly higher revenues and profit in data supply.

5.1.3 Scenario 1b_AC

The DDS strategy produces higher revenues and lower cost in scenario 1b_AC (see Fig. 6). Supplier A achieves the highest profit, €36.66, when the AQI data update time is 2 min and 52 s. In this case, the profit is €190.92 for data supplier C.

5.1.4 Scenario 1b_BC

Figure 6 shows that the DDS strategy produces the higher revenues and lower cost in scenario 1b_BC. The profit is

€142.41 for supplier C and €63.41 for supplier B. The difference in profit comes especially from cost savings, as the cost is €65.29 for supplier B and €9.09 for data supplier C.

5.1.5 Scenario 1c

The highest profit (or the lowest loss) for supplier A is −€23.57 when the AQI data update time is 2 min and 58 s (see scenario 1c in Fig. 6). In this case, the profit is €143.08 for supplier C and €55.85 for supplier B. Supplier C earns the highest revenues but also has the lowest cost, as the cost is €9.12 for supplier C, €26.17 for supplier A, and €69.55 for supplier B.

The results from the simulation scenario 1 show that the ODS$_{max}$ and DDS strategies assist achieving efficiency, dynamism, and flexibility in sellable data supply. Thus, these strategies optimize data trade and extend the work done in the prior studies presented in Table 1.

5.2 Results from simulation scenario 2

Simulation scenario 2 focuses on two data suppliers that supply high- and the low-quality situational AQI data for air quality change events. The AQI event time is in the range of 30 s to 15 min in the simulations. The suppliers use the following strategies in data supply:

(1) *Data supplier D* relies on EODS in supply of situational AQI data (DLDPs).
(2) *Data supplier E* relies on EDDS. The data supplier reacts to AQI change events, publishes showcase data products in the marketplace, and finally prepares the situational AQI data for data users' requests.

Scenario 2 consists of sub-scenarios 2a and 2b. In scenario 2a, there is only one supplier and no competition in data supply. In scenario 2b, there are two rivals in data supply. Figure 7 shows that the simulated profits, revenues, and costs decrease as the AQI event time increases. Data supplier D has lower cost, but data supplier E achieves clearly higher revenues and profits in scenarios 2a and 2b. For example, for a 1-min AQI event time, data supplier E's revenue is €3000, and cost is €180 in scenario 2a. In this case, the revenue is €2600, and cost is €129.58 for supplier D.

The results from the simulation scenario 2 show that the EDDS strategy assists achieving higher profits in situational AQI data supply, especially in the case of shorter AQI event time. Thus, the EDDS strategy improves efficiency, dynamism, and flexibility in sellable data supply and this way optimizes data trade and extends the work done in prior studies presented in Table 1.
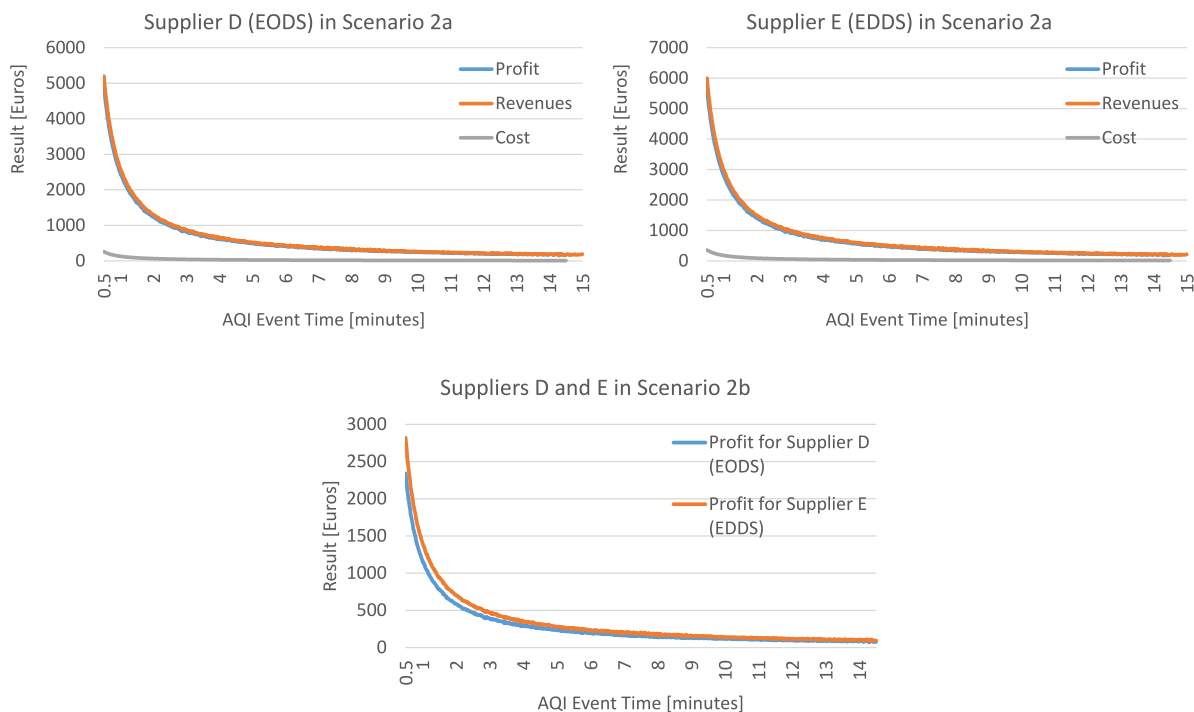

Fig. 7 Simulated profit for data supplier D and E as the AQI event time increases from 30 s to 15 min in scenarios 2a and 2b

6 Discussion and implications

The high-granularity data trade aims at increasing granularity as well as market, situational, and context awareness in data supply and use. The MSC-messages support this, and a layer that contains components for continuous data supply and data use optimization can be provided. However, there must be a fast network infrastructure and a marketplace that offers sufficient APIs for data exchange and trade. In addition, integration is needed to connect the layer to the marketplace and to the data use and supply components.

The data supply strategies have different strengths and weaknesses that must be considered when using these. For example, there are use case-specific requirements for the data supply strategies. ODS and EODS can serve instant use of data as DDS and EDDS enable preparation of data processing capabilities (e.g., edge computing capabilities) for potential data demand and adapting the data for data users' requirements. EODS and EDDS support supply of situation-dependent data. Data suppliers can use a set of strategies in data supply. For example, a data supplier can use the DDS strategy first and prepare the data after there is a user for it and subsequently use the ODS strategy in the selling of the prepared data for new data users, too.

DDS increased efficiency in data supply in the simulated scenarios. There was higher profit and revenues and lower costs in DDS than in ODS, and it seems that DDS is a better strategy if there is great fluctuation in data demand or only a small number of data buyers in the marketplace. However, DDS cannot serve use cases that require ready prepared data.

The ODS strategy with market share maximization (ODS_{max}) produced higher profits and revenues than the ODS strategy with periodic updates (ODS_p) in the simulated scenarios. However, the costs were higher in ODS_{max} than in ODS_p. EDDS increased efficiency in data supply in the simulated scenarios. The costs were lower in EODS, but EDDS produced clearly higher revenues and profits in the simulated scenarios.

Table 3 presents a SWOT-analysis of the ODS, DDS, EODS, and EDDS strategies.

ODS supports the supply of data for steady data market. These data can be statistical/historical data, slowly changing or periodically updated data, or data that do not require adaptation. The ODS and EODS strategies provide ready prepared sellable data, but there is a risk that there are no buyers for the data. In addition, the lack of input data can prevent preparation of sellable data in advance in ODS and EODS.

The DDS and EDDS strategies enable the supply of adapted and more up-to-date data for demand. The capabilities, such as edge computing capabilities, can be prepared for data supply, and the showcase data products make these capabilities visible in the marketplace. There can be a high volume of input data requiring complex processing and increasing latencies in demand-driven data supply or then a part of the data processing must be performed before there is a user for the data. For example, calculating an average for the last 5 min of AQIs requires continuous processing and storing of the AQI values as time-sequence data, or then it takes more than 5 min to collect the data and to calculate the average for the last 5 min of AQIs, and deliver the result for the data user.

7 Conclusion

The idea of the high-granularity data trade is to increase granularity as well as situational and context awareness in sellable data supply and use. This paper introduces four data supply strategies for high-granularity data trade. ODS and EODS offer fully prepared data that shortens processing delays in data use. However, there is a risk that there is no demand for the prepared data. The DDS and EDDS strategies decrease the unnecessary data preparation and enable adaptation of the data. However, the processing delays are higher than in ODS or in EODS, as the data are processed for data users' requests.

DDS increased efficiency in data supply in the simulated scenarios. There was higher profit and revenues and lower costs in DDS than in ODS. However, there are use cases that require the use of ODS, as DDS does not offer ready prepared data for instant use of data. EDDS increased efficiency in data supply in the simulated scenarios. The costs were lower in EODS, but EDDS produced clearly higher revenues and profits.

The MSC-messages assist in the data supply and data use optimization but cause additional network traffic in data trade. The full deployment of the strategies requires changes to the data supply and data use components and to data products. The data marketplace should offer API-call-based charging mechanisms for data products that enable charging users for data requests in DDS/EDDS. Second, there should be a mechanism for data product tailoring and metadata to determine tailoring options for data products. There can be extension components that add these capabilities to the marketplace, but it is required that the data marketplace APIs offer sufficient methods to enable the implementation of these components. ODS_{max} requires information about active data users in a marketplace. If the marketplace does not offer this information, the DUO and MD components and MSC-messages can be used for delivering this information for the data suppliers.

Table 3 SWOT Analysis of the ODS, DDS, EODS, and EDDS strategies

	ODS	DDS	EODS	EDDS
Strengths	• Sellable data is ready to be used	• Smaller risks and overhead in data supply as the data is prepared for demand • A data user can tailor data products before buying them • Data can be adapted for data users' requirements	• Sellable situational data is ready for instant use	• Supply of the latest situational data as the data is prepared for data users' requests • Supply of adapted situational data • Smaller risks and overhead in data supply as the data is prepared for demand
Weaknesses	• ODS can be used only when it is possible to prepare the sellable data beforehand • Data is not adapted for a data user, situation, and context	• More complex to implement than ODS • Instant use of data is not possible if it is slow to prepare the sellable data for use	• There is processing overhead if there are no buyers for the prepared data	• More complex to implement than EODS • Instant use of data is not possible if it is slow to prepare the sellable data for use
Opportunities	• Supply of data for steady data markets • Supply of statistical/historical, slowly changing, or periodically updating data • Supply of data that does not require adaptation	• Preparation of capabilities for data supply • Supply of tailorable data products and adapted data • Supply of small data volumes	• Supply of situational data for decision-making and monitoring • Increasing situational awareness in cities	• Preparation of capabilities for data supply • Supply of situational data for decision-making and monitoring • Increasing situational awareness in cities
Threats	• Lack of input data prevents data preparation in advance • There is a risk that there is no use and demand for the prepared data	• It is too difficult to integrate DDS features to data products and data supply components • Marketplace limits use of showcase data products in data trade and does not offer API-call-based pricing models for data products	• Lack of input data prevents preparation of supplementary data in advance • There is a risk that there is no use and demand for the prepared data	• It is too difficult to integrate EDDS features to data products and data supply components • Marketplace limits use of showcase data products in data trade and does not offer API-call-based pricing models for data products

The findings can have practical and social implications in smart cities. The high-granularity data trade and suggested data supply strategies assist optimization of sellable data supply for data demand and offer ways to improve cost and energy efficiency in sellable data supply, delivery, and use. This, in turn, can provide better sellable data for citizens, businesses and data-driven services that assist everyday life in cities.

Several actions were performed for ensuring the correctness of the simulation results. The simulator was developed incrementally and tested by performing very limited and simple simulation scenarios first. The goal was to detect the possible errors in the simulator, model, scenario, and instrumentation and fix these errors before performing the actual simulations. In addition, realistic simulation parameters were tried to be determined for the simulations. However, although simulated information was produced for the suggested data supply strategies, there is a still a clear need for further studies and simulations to improve the reliability and validity of the results. For example, there are limitations in the simulation method and uncertainties in the used simulation parameters (see Table 2) that affect the accuracy of the results. The following paragraphs discuss elements that affect the accuracy, reliability, and validity of the simulation results.

7.1 Market information

Data marketplaces are evolving but currently only a small part of available data is provided as sellable data products. In practice, this means that it is difficult to get real market information about the structure of markets, about data products' pricing, and about shares of different kinds of data users such as cost and quality-oriented users in the marketplaces. Without market information, it is also challenging to measure the reliability and validity of the simulation results. For example, the comparison information about the usability of the suggested data supply strategies in real use cases would assist the evaluation of the accuracy of the simulation results.

7.2 Simulation model

A very simplified model was used in the simulation scenarios that focused on a fixed set of data suppliers that do not change strategy in data supply. However, there can be great dynamism in the real-world data markets: new data suppliers can emerge who continuously optimize data supply and data pricing, and use different strategies in the data supply, affecting the data demand, too. The more comprehensive modeling and simulation of this dynamism would improve the accuracy of the results and offer information of the use of data supply strategies in changing data markets.

7.3 Simulation parameters

Although realistic simulation parameters were tried to be determined, there is a much uncertainty in these values as there are no measured data about these simulation parameters available (e.g., about data product prices for different quality-levels). The following paragraphs discuss the used simulation parameters and propose ways to improve the accuracy of the simulation results.

7.4 Earning and pricing parameters

The simulations require estimates for the earnings from a single data product (in the case of ODS and EODS) and from a single data request (in the case of DDS and EDDS). Although the realistic estimates were attempted to be determined, there is uncertainty in these estimates. Real market information of AQI data pricing would improve the estimates and assist in obtaining more accurate values for data suppliers' profits and revenues.

7.5 Cost parameters

The simulations require estimates for the input data use, data processing, and data product preparation costs. The errors in these estimates affect the simulated cost and profits of the data suppliers. Unfortunately, it is difficult to estimate these costs at a generic level. For example, the data processing cost depends on the used processing capabilities (e.g., the used cloud or edge-computing capabilities) and the input data cost can depend on the local data markets in a city.

7.6 Data user distribution

In simulations, the number of data users follows the hourly distribution of vehicle traffic that was measured in the city of Helsinki. However, it is challenging to estimate the share of cost-oriented users and quality-oriented users in a specific city or city district. Now, only a rough estimate for the user distribution is used, as there is the same number of users from each data user group in the simulations. Information of user distribution in the marketplaces and cities would minimize the errors in these estimates and assist in the planning and targeting of the data supply for demand.

In summation, market information and city-level information are needed to assist in the estimation of data pricing, data preparation costs, and data user distribution in a specific city or city district.

Appendixes

Appendix 1: Abbreviations

AQI	Air Quality Index
CAV	Connected and automated vehicle
CO	Carbon monoxide
DDS	Demand-driven data supply
DLDP	District-level data product
DPCs	Data and processing capabilities
DSN	Data supply notification
DSO	Data supply optimization
DUN	Data use notification
DUO	Data use optimization
EODS	Event and offering-driven data supply
EDDS	Event and demand-driven data supply
EDPC	Edge servers' data and processing capability
GDPR	General data protection regulation
IDS	International data spaces
IoT	Internet of Things
MD	Market driver
MQTT	Message queuing telemetry transport
MSC	Market, situation, and context update
NO_2	Nitrogen dioxide
O_3	Ozone
ODS	Offering-driven data supply
ODS_{max}	Offering-driven data supply with market share maximization
ODS_p	Offering-driven data supply with periodic updates for data offering
$PM_{2.5}$	Fine particles
PM_{10}	Respirable particles
SO_2	Sulfur dioxide
TRS	Total reduced sulfur compounds
ULDP	Ultra-local data product

Appendix 2: Formulas

The following formulas are used for calculating cumulative profit, revenues, and cost for each data supplier from the start to simulation time t:

(1) *Cumulative profit*: $Profit(t) = Revenues(t) - Costs(t)$

(2) *Cumulative revenues*:

$$Revenues(t) = \sum_{transaction \in Transactions(0,t)} Revenues(transaction)$$

In the formula, the *Transactions (0, t)* is a collection of the data supplier's data sale transactions from the start to the simulation time (t). The revenues from a data sale transaction depend on the used data supply strategy and are calculated as follows:

a. *Revenues from a data sale transaction in ODS and EODS*:

$$Revenues_{ODS\_and\_EODS}(transaction) = EarningsFromDataProduct$$

b. *Revenues from a data sale transaction in DDS and EDDS*:

$$Revenues_{DDS\_and\_EDDS}(transaction) = countOfDataRequest * DataRequestPrice$$

(3) *Cumulative cost* related to a data product from the start to the simulation time (t):

$$Cost(t) = \sum_{dataProduct \in DataProducts(0,t)} Cost(dataProduct, t)$$

The *DataProducts(0, t)* contains the data products that a data supplier has produced from the start to the simulation time (t). The cumulative data production cost depends on the data supply strategy and is calculated as follows:

a. *Data production cost in ODS and EODS*:

$$Cost_{ODS\_and\_EODS}(dataProduct, t) = InputDataCost + DataPreparationCost$$

b. *Data production cost in DDS and EDDS*:

$$Cost_{DDS\_and\_EDDS}(dataProduct, t) = DataRequestCount(dataProduct, t) * (InputDataCost + DataAdaptationCost)$$

Data availability No datasets were generated or analysed during the current study.

Declarations

References

Abbas AE, Ofe H, Zuiderwijk A, De Reuver M (2022) Preparing future business data sharing via a meta-platform for data marketplaces: exploring antecedents and consequences of data sovereignty. In: 35th Bled eConference–digital restructuring and human (re) action, pp 571–586

Alvsvåg R, Bokolo A, Petersen SA (2022) The role of a data marketplace for innovation and value-added services in smart and sustainable cities. Commun Comput Inf Sci 1585 CCIS:215–230. https://doi.org/10.1007/978-3-031-06668-9_16

Anthony B (2023) Decentralized brokered enabled ecosystem for data marketplace in smart cities towards a data sharing economy. Environ Syst Decis 43:453–471. https://doi.org/10.1007/s10669-023-09907-0

Anthony Jnr B, Petersen SA, Ahlers D, Krogstie J (2020) API deployment for big data management towards sustainable energy prosumption in smart cities—a layered architecture perspective. Int J Sustain Energy 39:263–289

Badii C, Bellini P, Cenni D et al (2017) Analysis and assessment of a knowledge based smart city architecture providing service APIs. Futur Gener Comput Syst 75:14–29. https://doi.org/10.1016/j.future.2017.05.001

Berry BJL (2008) Urbanization. In: Marzluff JM, Shulenberger E, Endlicher W et al (eds) Urban ecology: an international perspective on the interaction between humans and nature. Springer US, Boston, pp 25–48

Bornholdt H (2021) Towards citizen-centric marketplaces for urban sensed data. In: Zirpins C (ed) Communications in computer and information science. Springer, Cham, pp 140–150

Bouzeghoub M, Peralta V (2004) A framework for analysis of data freshness. Proc ACM SIGMOD Int Conf Manag Data 59–67. https://doi.org/10.1145/1012453.1012464

Dalle O (2012) On reproducibility and traceability of simulations. In: Proceedings—winter simulation conference. IEEE, pp 1–12

Di Vito D, Kanerva M, Järveläinen J et al (2020) Safe and sustainable design of composite smart poles for wireless technologies. Appl Sci 10:1–19. https://doi.org/10.3390/app10217594

Docherty I, Marsden G, Anable J (2018) The governance of smart mobility. Transp Res Part A Policy Pract 115:114–125. https://doi.org/10.1016/j.tra.2017.09.012

Duch-brown N, Martens B, Mueller-langer F (2017) The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01

Eggers G, Fondermann B, Maier B, et al (2020) GAIA-X: technical architecture

Elliott D, Keen W, Miao L (2019) Recent advances in connected and automated vehicles. J Traffic Transp Eng 6:109–131. https://doi.org/10.1016/j.jtte.2018.09.005

EU (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR)

EU (2022) Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)

European Commission (2021) Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts

European Commission (2022) Proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)

Fernandez RC, Subramaniam P, Franklin MJ (2020) Data market platforms: trading data assets to solve data problems. Proc VLDB Endow 13(11):1933–1947

Fielding RT (2000) Architectural styles and the design of network-based software architectures. University of California, Irvine

FMI (2023) Air Quality Index. In: Finnish Meteorol. Inst. https://en.ilmatieteenlaitos.fi/air-quality-index

Fruhwirth M, Rachinger M, Prlja E (2020) Discovering business models of data marketplaces. In: Proceedings of the 53rd Hawaii international conference on system sciences, pp 5738–5747

Gröger C (2021) There is no AI without data. Commun ACM 64:98–108. https://doi.org/10.1145/3448247

Hassan N, Yau K-LA, Wu C (2019) Edge computing in 5G: a review. IEEE Access 7:127276–127289. https://doi.org/10.1109/access.2019.2938534

Hayek FA (1944) The road to serfdom. University of Chicago Press, Chicago

IDSA (2019) IDS reference architecture, version 3.0

Karnouskos S, Da Silva PG, Ilic D (2012) Energy services for the smart grid city. IEEE Int Conf Digit Ecosyst Technol 1–6. https://doi.org/10.1109/DEST.2012.6227925

Khan Z, Kiani SL (2012) A Cloud-based architecture for citizen services in smart cities. In: Proceedings—2012 IEEE/ACM 5th international conference on utility and cloud computing, UCC 2012. IEEE, pp 315–320

Krejcar O, Maresova P, Selamat A et al (2019) Smart furniture as a component of a smart city-definition based on key technologies specification. IEEE Access 7:94822–94839. https://doi.org/10.1109/ACCESS.2019.2927778

Li B, Kisacikoglu MC, Liu C et al (2017) Big data analytics for electric vehicle integration in green smart cities. IEEE Commun Mag 55:19–25. https://doi.org/10.1109/MCOM.2017.1700133

Liang F, Yu W, An D et al (2018) A survey on big data market: pricing, trading and protection. IEEE Access 6:15132–15154. https://doi.org/10.1109/ACCESS.2018.2806881

Lilleberg I, Hellman T (2015) Liikenteen kehitys Helsingissä vuonna 2010

Liu Z, Hacigümüş H (2014) Online optimization and fair costing for dynamic data sharing in a cloud data market. In: Proceedings of the ACM SIGMOD international conference on management of data, pp 1359–1370

Lorenzo B, Gonzalez-Castano FJ (2016) A matching game for data trading in operator-supervised user-provided networks. In: 2016 IEEE International conference on communications, ICC 2016. IEEE, pp 1–7

Maier MW (1998) Architecting principles for systems-of-systems. Syst Eng 1:267–284. https://doi.org/10.1002/(SICI)1520-6858(1998)1:4%3c267::AID-SYS3%3e3.0.CO;2-D

Mangiuc DM (2012) cloud identity and access management—a model proposal. Account Manag Inf Syst Contab Si Inform Gestiune 11:484–500

McGrath H, Kotsollaris M, Stefanakis E, Nastev M (2019) Flood damage calculations via a RESTful API. Int J Disaster Risk Reduct 35:101071. https://doi.org/10.1016/j.ijdrr.2019.101071

Palviainen M, Kotovirta V (2021) Approach for simulating vehicle-based supply of sellable data products in smart cities—parking space data as a use case. J Urban Mobil 2:22

Palviainen M, Suksi J (2023) Data marketplace research: a review of the state-of-the-art with a focus on smart cities and on edge data exchange and trade. In: 2023 Smart cities symposium Prague (SCSP), pp 1–7

Park S (2017) OCF: A new open IoT consortium. In: 31st IEEE International conference on advanced information networking and applications workshops, WAINA 2017. IEEE, pp 356–359

Patti E, Acquaviva A (2016) IoT platform for smart cities: requirements and implementation case studies. In: 2016 IEEE 2nd Int Forum Res Technol Soc Ind Leveraging a Better Tomorrow, RTSI 2016, pp 1–6. https://doi.org/10.1109/RTSI.2016.7740618

Ramachandran GS, Radhakrishnan R, Krishnamachari B (2018) Towards a decentralized data marketplace for smart cities. In: 2018 IEEE International smart cities conference, ISC2 2018. IEEE, pp 1–8

Robbins DE, Tanik MM (2014) Cyber-physical ecosystems: app-centric software ecosystems in cyber-physical environments. In: Suh S, Tanik U, Carbone J, Eroglu A (eds) Applied cyber-physical systems. Springer, New York, pp 141–147

Robin M, Botts A (2007) OpenGIS sensor model language (SensorML) implementation specification, OGC 07-000

Sharma P, Lawrenz S, Rausch A (2020) Towards trustworthy and independent data marketplaces. In: ACM International conference proceeding series. pp 39–45

Shi W, Cao J, Zhang Q et al (2016) Edge computing: vision and challenges. IEEE Internet Things J 3:637–646. https://doi.org/10.1109/JIOT.2016.2579198

Simmhan Y, Ravindra P, Chaturvedi S et al (2018) Towards a data-driven IoT software architecture for smart city utilities. Softw Pract Exp 48:1390–1416. https://doi.org/10.1002/spe.2580

Spiekermann M (2019) Data marketplaces: trends and monetisation of data goods. Intereconomics 54:208–216. https://doi.org/10.1007/s10272-019-0826-z

Studley ME, Little H (2021) Robots in smart cities. How smart is your city? Intell Syst Control Autom Sci Eng 98:75–88. https://doi.org/10.1007/978-3-030-56926-6_7

Taylor L, Mukiri-Smith H, Petročnik T et al (2022) (Re)making data markets: an exploration of the regulatory challenges. Law Innov Technol 14:355–394. https://doi.org/10.1080/17579961.2022.2113671

Tiddi I, Bastianelli E, Daga E et al (2020) Robot-city interaction: mapping the research landscape—a survey of the interactions between robots and modern cities. Int J Soc Robot 12:299–324. https://doi.org/10.1007/s12369-019-00534-x

Tokuda H (2004) Smart furniture: a platform for creating context-aware ubiquitous applications everywhere. In: EUC 2004. LNCS, p 1112

UN (2022) The sustainable development goals report 2022

Verma A, Prakash S, Srivastava V et al (2019) Sensing, controlling, and IoT infrastructure in smart building: a review. IEEE Sens J 19:9036–9046. https://doi.org/10.1109/JSEN.2019.2922409

Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 12:5–33

Xu X, Liu LQ, Zhang L et al (2020) Abnormal visual event detection based on multi-instance learning and autoregressive integrated moving average model in edge-based smart city surveillance. Softw Pract Exp 50:476–488. https://doi.org/10.1002/spe.2701

Yerabolu S, Kim S, Gomena S et al (2019) DeepMarket: an edge computing marketplace with distributed TensorFlow execution capability. In: IEEE INFOCOM 2019-IEEE conference on computer communications workshops. IEEE, pp 32–37

Zheng Z, Xie S, Dai HN et al (2020) An overview on smart contracts: challenges, advances and platforms. Futur Gener Comput Syst 105:475–491. https://doi.org/10.1016/j.future.2019.12.019