Responsible Sharing of Spatiotemporal Data

Raul Castro Fernandez
The University of Chicago
USA
raulcf@uchicago.edu

Arnab Nandi
The Ohio State University
USA
nandi.9@osu.edu

ABSTRACT

There is a growing need for responsible spatiotemporal data sharing in our daily lives. Applications such as connected vehicles and mobile advertising are currently undergoing a significant digital shift, demanding new standards for privacy-aware solutions and the integration of machine learning technologies. In this tutorial, we present the concepts and challenges encountered when maximizing the utility of spatiotemporal data while enforcing rigorous privacy and security measures. We review modern data sharing mechanisms that provide stakeholders with the power to establish precise terms for the usage and sharing of their data, secured by a robust data infrastructure. We will explore how such sharing mechanisms interplay with complex privacy stipulations and advanced spatiotemporal analytics. Attendees will leave with a comprehensive understanding of how to navigate the delicate balance of spatiotemporal data usage, paving the way for innovation in privacy and compliance methodologies across various industries.

CCS CONCEPTS

• **Information systems** → *Data exchange*; *Spatial-temporal systems.*

KEYWORDS

Responsible Data Sharing, Spatiotemporal, Data Exchange

ACM Reference Format:
Raul Castro Fernandez and Arnab Nandi. 2024. Responsible Sharing of Spatiotemporal Data. In *Companion of the 2024 International Conference on Management of Data (SIGMOD-Companion '24), June 9–15, 2024, Santiago, AA, Chile.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3626246.3654688

1 INTRODUCTION

Spatiotemporal data [1] is characterized by its spatial and temporal attributes. It is a major source of value across applications and sectors, including but not limited to mobile phone advertising, vehicular telematics, urban planning, environmental monitoring, public health, and logistics. Sharing spatiotemporal data generates value for all these applications, but it is challenging due to the sensitiveness of the data. Spatiotemporal data reveals fine-grained geographical locations over specific time periods, introducing important challenges related to privacy and compliance. This tutorial's objective is to lay out the principles of spatiotemporal data sharing, discuss privacy implications of such data, and describe concrete techniques used to responsibly handle this kind of data.

As an example, modern vehicles are equipped with 3G/4G/LTE/5G-based telematics capabilities that transmit vehicle location, trajectory, and on-board data to power mobility services, help improve driver safety, inform infrastructure operations, and more. These streams of data are valuable by themselves. But when combined with other spatiotemporal data they become even more valuable [25]; e.g., better precision in smart city use cases such as dynamic intersections and citizen reporting of infrastructure issues, or commercial offerings such as usage-based insurance, risk management [14, 15], and en-route recommendations. Automakers who otherwise with each other could simultaneously benefit from collaboration over this data to improve driver safety.

Despite the upsides, sharing such fine-grained spatiotemporal data is a critical privacy and security concern e.g., if data is misused or rogue actors overshare. A recent report has highlighted the potential for damage [19] and has motivated legislators to question automakers about their spatiotemporal data sharing practices [6]. While mitigating these issues is paramount, it also can risk leaving the potential value of *responsible* sharing untapped. Data *flows* from individual vehicles to the automakers' (or third-party telematics service providers') infrastructure. However, collection and retention of such data are at odds with privacy, security, and compliance concerns, which often result in unmaterialized dataflows. Data is simultaneously valuable and a liability. The threat of losing control of dataflows and leaking data beyond what's permitted is a significant impediment to unleashing the power of data science over such rich pooled datasets, and this is not only true in the automotive telematics space but across sectors and applications.

What is covered: This tutorial will provide a structured overview of spatiotemporal data management, focusing on the challenges and techniques for secure, efficient, and compliant data sharing. It will cover the fundamentals of spatiotemporal data characteristics, delve into privacy and security concerns, and explore responsible dataflow management strategies. Participants will learn about implementing privacy-preserving protocols and frameworks for data sharing, with practical use cases such as connected vehicles. Additionally, the tutorial will address future trends, emerging technologies, and potential challenges in the field, equipping attendees with theoretical knowledge that will help them consider how to balance the pros and cons of sharing spatiotemporal data.

Intended Audience

We expect the audience to be data management researchers, industry professionals in sectors reliant on spatial and temporal data

SIGMOD-Companion '24, June 9–15, 2024, Santiago, AA, Chile

ACM ISBN 979-8-4007-0422-2/24/06
https://doi.org/10.1145/3626246.3654688

data, and academics including faculty and graduate-level students. Attendees are expected to have a foundational understanding of database concepts and systems, as well as a basic familiarity with issues surrounding data privacy and security. Prior exposure to spatial or geographic information systems (GIS) would be beneficial but not necessary; we will design the tutorial to to equip participants with the knowledge necessary to engage with cutting-edge research and applications in spatiotemporal data sharing.

2 MOTIVATION

Our tutorial is motivated by the rapid growth and use of spatiotemporal data in multiple industries. Modern smartphones are equipped with location sensing capabilities using capabilities such as GPS and cell tower triangulation, powering industries such as advertising at the billion-person scale [10]. Simultaneously, there are expected to be 470 million connected vehicles by 2025, representing a global industry worth hundreds of billions. Cars' sensors capturing driver behavior and routes pave the way to answer questions related to safety, urban analytics, advanced driver assistance systems (ADAS), self-driving, and logistics. At a time when the automotive industry is going through a digital transformation prompted by the introduction of electric vehicles and advanced telematics sensors, it is more important than ever to provide responsible infrastructure to collect, combine, and analyze data to extract its value while controlling for privacy and compliance risks of highly sensitive data. There are similar trends of growth of spatiotemporal data in domains such as smart cities [17], aviation [16], and public health. Together, these depict the contours of a large opportunity locked behind important challenges. The tutorial will explain the main use cases around spatiotemporal data sharing as well as the principles and technologies in use today to enable such a process. We pay special attention to the compliance, privacy challenges:

Transparent Privacy Compliance: We will introduce techniques and protocols to enable data scientists work across shared datasets without worrying about inadvertent violations of privacy regulations, and ensure that they are compliant with all listed regulations.

Improved Collaborations: We will discuss techniques to establish collaborations among participants. These techniques are geared toward reducing the logistics involved in forming data sharing partnerships and the risks of data sharing.

Reduced Privacy-related Backtracking: When working outside controlled environments, data scientists may face a risk of privacy violations that may require re-analysis or data disposal, due to misspecification of privacy preferences, or implementation issues. Due to the controlled setup, the impact of such incidents is mitigated.

Here, we review ideas and tooling from both academia and industry. On the academic front, we will discuss privacy-enhancing technologies (*PETs*) such as multi-party computation [8], and data escrow systems [26]. On the industrial front, we will offer an overview of data clean room technologies [7] and privacy and confidentiality-focused data infrastructure as offered by major cloud providers and platforms such as Microsoft Azure AE, Snowflake, Databricks, Microsoft CCF [20], and PySyft [31]. Our goal is to present the capabilities of these different techniques to the audience and trace those back to the requirements of practical data sharing scenarios.

3 CONCEPTS COVERED

The tutorial is divided into two parts, starting with a primer covering the fundamentals of spatiotemporal data and data sharing. The second part presents use cases to illustrate ideas from the first part.

3.1 Fundamentals

Spatiotemporal Data and Primitives: We formally define spatiotemporal data. We introduce the basic primitives and publicly available datasets to explore this data, offering illustrative examples. Then, we look into *sourcing* such data – processes and technologies involved in collecting, processing, and analyzing large volumes of data from different sources – including vehicles – which include their location and trajectory over time. The pipelines are designed to leverage this data for various applications, such as smart transportation systems, safety features, and enhanced mobility services.

Spatial and Spatiotemporal Joins: Spatial joins [9, 27] allow spatial datasets to be merged based on geographical relationships, such as location, proximity, or overlap. For example, this method often pairs data points from one set with their nearest neighbors in another, based on spatial criteria, enabling the efficient integration and analysis of disparate geospatial data sources. On the other hand, a *spatiotemporal* [1, 5] join expands upon this concept by also incorporating the element of time, merging datasets not just based on spatial proximity but also on temporal alignment. This type of join considers both the location and the time frame of the data points, facilitating a more dynamic analysis that is crucial in tracking movements over time, making them indispensable in fields where both spatial and temporal factors are crucial, such as vehicle telematics, urban planning, and environmental monitoring.

Controlling Dataflows: We define dataflow and present the dataflows that materialize in several spatiotemporal scenarios. Controlling dataflows [4] refers to the secure and regulated transfer of telematics data between stakeholders, e.g., vehicles, automakers, and potentially other third parties. It ensures that the data is only accessed and used by authorized entities in a manner that is compliant with privacy regulations and security protocols.

Privacy & Regulatory frameworks, Compliance Contracts: Frameworks like the GDPR (General Data Protection Regulation) [13, 24] in the European Union and the CCPA (California Consumer Privacy Act) in California play a crucial role in managing and protecting spatiotemporal data, such as that used in telematics. These laws, and the underlying frameworks, were enacted to address the complexities and risks associated with the collection, processing, and storage of personal data, particularly in the context of advanced digital technologies and the global flow of information. These frameworks set strict guidelines on how personal data, including location and time-specific information, should be collected, processed, and stored. They emphasize the need for consent from individuals before their data is used, ensure transparency in data handling processes, and mandate prompt notifications in case of data breaches. For businesses and organizations dealing with telematics data, compliance with these regulations is essential not only for legal conformity but also for maintaining user trust, managing expectations, and protecting individual privacy. This is particularly

relevant in a world where data is increasingly used for a multitude of applications, from improving transportation systems to personalized marketing, making the ethical and secure handling of such information a top priority.

We will further discuss examples of contracts or agreements that outline how data must be handled to comply with the above privacy laws and regulatory frameworks. These contracts define the obligations of parties to protect sensitive information and establish the parameters for data usage and sharing.

Data Exchanges, Escrows, and Pools: There is a growing need for data *exchanges* – platforms or mechanisms that facilitate the secure exchange of vehicle data between parties, allowing for the data to be shared and combined with other datasets to enhance its value, while also ensuring compliance with privacy and security requirements. In this setting, the placement of data can often be a critical concern, especially when considering analysis or computation over multi-party data. Here, a *data escrow* can be implemented: trustworthy intermediaries [26] that hold and protect vehicle data and only release it when predefined conditions are met. These conditions would be aligned with privacy and security requirements, ensuring data is used appropriately and reducing the liability for data providers. Another factor to consider is that combining data across multiple parties increases its value beyond the sum of the parts. In this part of the tutorial we explain how, and present the principles of *data pooling*, including incentives [4, 30], compliance, and technical challenges to achieve such pools.

Query Federation: One approach to dealing with private data is to federate the query itself. Systems such as HUFU [17] optimize federated spatial queries like range, counting, and kNN join, ensuring high efficiency and usability, while also decomposing queries into plaintext and secure operators, balancing performance and security. Such a federation approach differs from data escrows and other Privacy Enhancing Technologies (PETs) in several key ways. While data escrows centralize data storage and manage access through a trusted third party, spatial query federation allows data to remain distributed across multiple owners. The distributed coordination in federated systems versus centralized systems such as data escrows pose tradeoffs in security, efficiency, and management overhead, which we will walk through in the session.

Federated Learning: Federated learning [29] offers a host of techniques to train a *central* model while keep data *locally*. If moving data is a problem, federated learning provides a technical solution to avoid that—even though some leakage still takes place through the exchanged gradients [28]. Furthermore, the techniques are generally limited to training certain machine learning models. Thus, we leave an in-depth discussion of this line of work out of the tutorial.

3.2 Spatiotemporal Data Sharing Use Cases

We will illustrate use cases for spatiotemporal data sharing in domains such as ad tech and vehicular telematics. The ad-tech business depends on combining data from multiple parties privately, e.g., so that a vendor understands the performance of an advertising campaign per geographic area. Vehicles are often called "smartphones on wheels", in other words, massive sources of data streams valuable to both drivers and society – when such data leads to bettering transportation systems and overall safety. We overview the challenges of combining vehicular data and the value that arises from such combination. Through a real-world use case, we illustrate the principles outlined in the first part of the tutorial. As part of our use case, we overview the different data streams vehicles produce, both those that are transmitted in real-time to the cloud and those that, while remaining in the vehicle while driving, are eventually shared. We overview the origins of vehicle data, which in this context may include individual vehicles, fleets, and telematics providers. Multiple stakeholders such as automakers, government agencies, and third-party vendors have an interest in accessing and using this data for various purposes. Specifically, we highlight the challenges of integrating spatiotemporal data with different granularities, and the data quality and robustness challenges that arises from such a process. We then explain mechanisms to simplify the problem and practical solutions to gain value from the combined data.

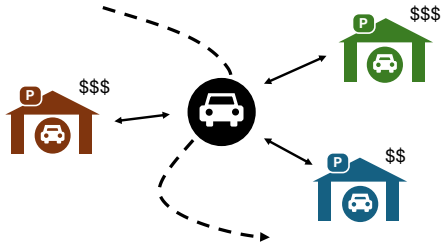

Figure 1: Example Use Case: Parking Recommendations can be based on factors such as proximity, price, and availability, but also consider the privacy of the vehicle's trajectory.

Example use case: Parking Recommendations: To illustrate the concepts in this tutorial, we will walk through an example use case of a parking recommendation system available in some vehicle's dashboards. City parking is a large-scale problem globally, e.g., parking in Chicago generated over $150 million in 2021 [23]. Such parking recommendation systems combines the vehicle's real-time location, obtained through GPS and telematics, with live parking availability data from nearby garages. This integration allows the system to continuously monitor both the car's current position and the occupancy status of various parking facilities. By processing this combined data, the system can identify optimal parking options based on factors like distance, availability, and user preferences, such as price. Recommendations are dynamically updated and displayed to the driver, often through the vehicle's infotainment system, complete with navigational directions to the chosen parking spot. This approach not only streamlines the parking process for the driver but also contributes to better traffic flow in urban areas by reducing the time vehicles spend searching for parking.

Integrating a car's future planned trajectory into a parking recommendation system enhances its precision but significantly heightens privacy concerns. This feature involves sensitive data that not only reveals the driver's current location but also possibly their intended destinations and travel patterns, posing risks of unauthorized access and potential misuse. This data is inherently both spatial *and* temporal – both the vehicles location and parking availability are changing constantly. Furthermore, it highlights the

challenges of explicit user consent from drivers, and the need to adhere to privacy regulations like GDPR and CCPA. These measures are crucial to protect individual privacy, maintain user trust, and comply with legal standards while leveraging the advanced capabilities of spatiotemporal data in enhancing parking solutions.

In the tutorial, we will walk through how such a system can be implemented using spatial data escrows, a solution that combines the concepts described in the previous section. The framework, particularly pertinent to systems that integrate a vehicle's planned trajectory with parking recommendation services, serves as an intermediary, managing the access and use of sensitive data under well-defined conditions. This approach ensures that the spatial data, including future route information, is handled responsibly, balancing the enhanced functionality of telematics services with the crucial need for data privacy and security.

Here, we will first explore how a spatial data escrow provides controlled access to the trajectory data. It establishes a protocol where data is only released or used if certain criteria, such as user consent, compliance with privacy laws, and adherence to specific contractual terms, are met. This mechanism significantly minimizes the risk of unauthorized data access and misuse. We will delve into the methods of data anonymization and aggregation within the escrow, demonstrating how it can effectively mask individual data points to prevent the identification of specific users or vehicles, thereby aligning with privacy regulations like GDPR and CCPA.

Furthermore, we will cover a data escrow system's auditing and accountability features, which provide an essential layer of transparency and trust. By keeping detailed records of when and how the data is used, the escrow system not only facilitates compliance with regulatory requirements but also serves as a tool for resolving disputes and investigating potential breaches. Lastly, we will address how the spatial data escrow framework acts as a risk mitigation tool, ensuring that the data flows align with the predetermined rules and policies, thereby significantly reducing the liabilities associated with handling sensitive spatiotemporal data. This comprehensive overview will equip participants with a clear understanding of implementing and managing a spatial data escrow system, an indispensable component in the landscape of modern data sharing and analytics.

4 STRUCTURE OF TUTORIAL

The tutorial will have the following format:

Introduction to Spatiotemporal Data:
- Overview of Spatiotemporal Data: Definitions and Characteristics
- Working with GPS Data, Trajectories, and Aggregates
- Use cases in Spatiotemporal Data Management (e.g., Telematics, Advertising, Parking Example)

Basics of Data Privacy and Security
- Fundamentals of Data Privacy: Concepts and Principles
- Introduction to Data Security in Databases
- Regulatory Landscape: GDPR, CCPA, and more

Dataflow Management in Spatiotemporal Context
- Understanding Dataflows for Spatiotemporal Data
- Principles of Dataflow Design and Management
- Case Studies: Effective Dataflow Management in Practice

Privacy-Preserving Techniques in Data Sharing
- Techniques for Ensuring Privacy in Data Sharing
- Anonymization, Encryption, and Differential Privacy
- Practical Challenges and Solutions in Privacy Preservation

Advanced Data Sharing Mechanisms
- Designing Data Sharing Protocols and Frameworks
- Dataflow Contracts, Escrows, and Unions
- Role of Data Escrow Systems in Spatiotemporal Data

Survey of Existing Tools
- Tools for Spatiotemporal Data Exploration
- Tools for Data Sharing

Example Use Case Walkthrough
- Use Case: Parking Recommendations *(provided as Jupyter Notebook to follow along)*

5 OPEN PROBLEMS

Despite tremendous production and research-facing activities in this space, we are left with several open research questions, which we encourage the community to consider:

Complex Analytics Primitives: Given the complexity of privacy semantics, what is the impact of controlled dataflows on complex analytics primitives such as scalable aggregation techniques, indexing, and data compression?

Data Quality Challenges: Spatiotemporal data from sensor-sourced vehicles is often incomplete and suffers from several data quality issues; how do we build query platforms that simultaneously handle and represent uncertain, incomplete, and privacy-controlled data?

Usability and User Interaction: In terms of the end-users, what challenges and complexities does such a platform and controlled dataflow layer add, both from a systems perspective (e.g., latency), and from a user perspective (e.g, increased cognitive load)? How do we design spatiotemporal primitives that cleanly allow for expressibility and explainability [2, 21] without increasing complexity?

Data stewardship and Ethics: Even in the presence of performant and semantically correct systems, it is important to consider the ethical implications [11] of data sharing, including consent, stewardship of data, and potentials for misuse.

Systems Performance: Given the real-world performance requirements imposed by large volumes of spatiotemporal data, how do we implement data processing over encrypted data while maintaining high performance? For this, we consider the real-world performance requirements of such a system, with the need to support high-frequency, low-latency data fusion across hundreds of millions of vehicles.

Answers to these questions are crucial to enable trustworthy computation and readily provided by the data escrow infrastructure on top of which we build. Longer term, we hope that this introduction

of principles and mechanisms to facilitate data sharing while avoiding oversharing through this tutorial will spark the building of a larger community in this space.

6 PRESENTER BIOGRAPHIES

Raul Castro Fernandez's research focuses on data markets, data privacy, and building systems that aid in managing, sharing, and integrating private data. This includes Data Station [26], a data platform that enables governance and enforcement of complex data access policies, including the ability to pool data across users to build mutually beneficial machine learning models.

Arnab Nandi's research focuses on human-in-the-loop data infrastructure, including large-scale interactive analytics. Most recently, Nandi spun off his work on interactive spatiotemporal analytics into a startup which focused on analyzing connected vehicle data. Following its acquisition by Azuga Inc (now a Bridgestone company), Nandi served there as Vice President of Data Science.

REFERENCES

[1] Md Mahbub Alam, Luis Torgo, and Albert Bifet. 2022. A survey on spatio-temporal data analytics systems. *Comput. Surveys* 54, 10s (2022), 1–38.

[2] Julio Angulo, Simone Fischer-Hübner, Erik Wästlund, and Tobias Pulls. 2012. Towards usable privacy policy display and management. *Information Management & Computer Security* 20, 1 (2012), 4–17.

[3] Dan Boneh, Craig Gentry, Shai Halevi, Frank Wang, and David J Wu. 2013. Private database queries using somewhat homomorphic encryption. In *ACNS*. Springer.

[4] Raul Castro Fernandez. 2023. Data-Sharing Markets: Model, Protocol, and Algorithms to Incentivize the Formation of Data-Sharing Consortia. *Proceedings of the ACM on Management of Data* 1, 2 (2023), 1–25.

[5] Lisi Chen, Shuo Shang, Christian S Jensen, Bin Yao, and Panos Kalnis. 2020. Parallel semantic trajectory similarity join. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 997–1008.

[6] Ed Markey 2023. Senator Markey Queries 14 Automakers About Invasive Data Practices, Calls for Protections for Consumer Privacy in Vehicles. https://www.markey.senate.gov/news/press-releases/senator-markey-queries-14-automakers-about-invasive-data-practices-calls-for-protections-for-consumer-privacy-in-vehicles.

[7] iab Tech Labs 2023. Data Clean Rooms. https://iabtechlab.com/datacleanrooms/.

[8] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. 2009. Zero-knowledge proofs from secure multiparty computation. *SIAM J. Comput.* (2009).

[9] Edwin H Jacox and Hanan Samet. 2007. Spatial join techniques. *ACM Transactions on Database Systems (TODS)* (2007).

[10] Jon Keegan and Alfred Ng. 2023. There's a Multibillion-Dollar Market for Your Phone's Location Data. https://themarkup.org/privacy/2021/09/30/theres-a-multibillion-dollar-market-for-your-phones-location-data.

[11] Rob Kitchin. [n. d.]. The ethics of smart cities and urban science. *Philosophical transactions of the royal society* ([n. d.]).

[12] Agnieszka Leszczynski. 2015. Spatial big data and anxieties of control. *Environment and Planning D: Society and Space* 33, 6 (2015), 965–984.

[13] Kevin McDonnell, Finbarr Murphy, Barry Sheehan, Leandro Masello, German Castignani, and Cian Ryan. 2021. Regulatory and technical constraints: An overview of the technical possibilities and regulatory limitations of vehicle telematic data. *Sensors* 21, 10 (2021), 3517.

[14] Sobhan Moosavi, Behrooz Omidvar-Tehrani, R Bruce Craig, Arnab Nandi, and Rajiv Ramnath. 2017. Characterizing driving context from driver behavior. *SIGSPATIAL 2017* (2017).

[15] Sobhan Moosavi, Mohammad Hossein Samavatian, Arnab Nandi, Srinivasan Parthasarathy, and Rajiv Ramnath. 2019. Short and long-term pattern discovery over large-scale geo-spatiotemporal data. In *ACM SIGKDD*.

[16] Behrooz Omidvar-Tehrani, Arnab Nandi, Nicholas Meyer, Dalton Flanagan, and Seth Young. 2017. Dv8: Interactive analysis of aviation data. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE, 1411–1412.

[17] Xuchen Pan, Yongxin Tong, Chunbo Xue, Zimu Zhou, Junping Du, Yuxiang Zeng, Yexuan Shi, Xiaofei Zhang, Lei Chen, Yi Xu, et al. 2022. Hu-fu: a data federation system for secure spatial queries. *Proceedings of VLDB* 15, 12 (2022), 3582–3585.

[18] Franz Papst, Naomi Stricker, Rahim Entezari, and Olga Saukh. 2022. To Share or Not to Share: On Location Privacy in IoT Sensor Data. In *IoTDI*. IEEE, 128–140.

[19] Privacy Not Included 2023. It's Official: Cars Are the Worst Product Category We Have Ever Reviewed for Privacy. https://foundation.mozilla.org/en/privacynotincluded/articles/its-official-cars-are-the-worst-product-category-we-have-ever-reviewed-for-privacy/.

[20] Mark Russinovich, Edward Ashton, Christine Avanessians, Miguel Castro, Amaury Chamayou, Sylvan Clebsch, Manuel Costa, Cédric Fournet, Matthew Kerner, Sid Krishna, et al. 2019. CCF: A framework for building confidential verifiable replicated services. *Technical Report MSR-TR-201916* (2019).

[21] Norman Sadeh, Alessandro Acquisti, Travis D Breaux, Lorrie Faith Cranor, Aleecia M McDonald, Joel R Reidenberg, Noah A Smith, Fei Liu, N Cameron Russell, Florian Schaub, et al. 2013. The usable privacy policy project. In *Technical report, Technical Report, CMU-ISR-13-119*. Carnegie Mellon University.

[22] Mohamed Sarwat and Arnab Nandi. 2017. On designing a geoviz-aware database system-challenges and opportunities. In *Advances in Spatial and Temporal Databases: SSTD*. Springer International Publishing.

[23] Fran Spielman. 2023. Parking meter deal keeps on giving. https://chicago.suntimes.com/city-hall/2023/6/11/23755615/chicago-parking-meters-annual-audit-record-revenue.

[24] Félicien Vallet. 2019. GDPR and Its Application in Connected Vehicles—Compliance. In *CESA Automotive Electronics Congress*. Springer.

[25] Glenn Vancauwenberghe, Ezra Dessers, Joep Crompvoets, and Danny Vandenbroucke. 2014. Realizing data sharing: The role of spatial data infrastructures. *Open Government: Opportunities and Challenges for Public Governance* (2014).

[26] Siyuan Xia, Zhiru Zhu, Chris Zhu, Jinjin Zhao, Kyle Chard, Aaron J Elmore, Ian Foster, Michael Franklin, Sanjay Krishnan, and Raul Castro Fernandez. 2022. Data station: delegated, trustworthy, and auditable computation to enable data-sharing consortia with a data escrow. *Proceedings of VLDB* (2022).

[27] Jia Yu, Zongsi Zhang, and Mohamed Sarwat. 2019. Spatial data management in apache spark: the geospark perspective and beyond. *GeoInformatica* (2019).

[28] Kai Yue, Richeng Jin, Chau-Wai Wong, Dror Baron, and Huaiyu Dai. 2023. Gradient obfuscation gives a false sense of security in federated learning. In *32nd USENIX Security Symposium (USENIX Security 23)*. 6381–6398.

[29] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. 2021. A survey on federated learning. *Knowledge-Based Systems* 216 (2021), 106775.

[30] Pan Zhou, Qian Lin, Dumitrel Loghin, Beng Chin Ooi, Yuncheng Wu, and Hongfang Yu. 2023. Incentive-Aware Decentralized Data Collaboration. *Proceedings of ACM SIGMOD* (2023). https://dl.acm.org/doi/10.1145/3589303

[31] Alexander Ziller, Andrew Trask, Antonio Lopardo, Benjamin Szymkow, Bobby Wagner, Emma Bluemke, Jean-Mickael Nounahon, et al. 2021. Pysyft: A library for easy federated learning. *Federated Learning Systems: Towards Next-Generation AI* (2021), 111–139.