Letter from the Special Issue Editor

Software systems that store and process personal data have become ubiquitous over the last years and have enabled numerous economic, technological and scientific advances. Unfortunately, the benefits of data-driven analysis and decision making have also been accompanied by several negative developments. Examples include the increased surveillance capabilities of the state [3] and private companies [4], negative impact on economic inequality [2] and traumatic experiences for individuals [5]. As a reaction, many countries have started to regulate data storage and processing to guarantee and protect the rights of individuals. The most comprehensive such regulation is the General Data Protection Regulation (GDPR, `https://gdpr.eu`) issued by the European Union.

In this special issue on *Directions Towards GDPR-Compliant Data Systems and Applications*, we continue the ongoing discussion in the data management community [1] on how to redesign data systems and applications to be compliant with such regulation.

Data deletion as a first-class-citizen. The first three papers of this issue address an important question originating from the "right-to-be-forgotten" postulated by GDPR: *How can we design efficient data systems that support the timely deletion of data as a first-class citizen?* The first paper on *Disposal by Design* presents a vision for automating data disposal which takes into account processing constraints, regulatory constraints as well as storage constraints, and gives concrete examples from the e-commerce domain, including a suggestion of how to to find summaries of relational data with machine learning. The second paper on *Building Deletion-Compliant Data Systems* argues that the the requirement of timely deletion of user data is becoming central in modern data management scenarios. The authors present a new framework for building deletion-compliant data systems from a holistic perspective, analyse the requirements derived from the new policies, and propose changes in the application and the system layer of data management systems. The third paper called *Provenance-based Model Maintenance: Implications for Privacy* focuses on efficient data deletion in a machine learning context. In particular, the authors focus on the extremely challenging problem to refresh existing models after the removal of training samples, which is called "machine unlearning". They argue that GDPR regulations imply that the removed samples must be fully erased from the models so that they cannot be leaked to an adversary. The paper reviews two provenance-based solutions and shows how they can guard against "model inversion attacks", which reconstruct the removed training samples from the updated models after the unlearning process.

Efficient data processing under regulatory constraints. The subsequent two papers of this issue adress an orthogonal systems-related question originating from GDPR: *How can we design efficient data systems that comply with data processing regulations?* The fourth paper of this issue on *Navigating Compliance with Data Transfers in Federated Data Processing* presents work on novel systems and methods for federated data processing, where the processing of geo-distributed data is subjected to data transfer regulations. The authors showcase recent work on compliant geo-distributed data processing and present research challenges and opportunities in making federated data processing systems GDPR-compliant. The fifth paper called *Towards Privacy by Design for Data with STRM privacy* discusses the practical challenges of engineering teams to balance privacy and innovation, with respect to effort, data utility and computation costs. The authors argue that current approaches in scalable data systems often treat privacy as an access problem, which is at odds with important legal and design principles. Instead, the authors propose that engineering teams should shift their data privacy efforts to the point of data collection, and discuss an architectural setup for privacy-compliant stream processing applications, which is in production usage.

This work was supported by Ahold Delhaize. All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their respective employers and/or sponsors.

References

[1] S. Supreeth, V. Banakar, M. Wasserman, A. Kumar, and V. Chidambaram. Understanding and Benchmarking the Impact of GDPR on Database Systems. Proceedings of the VLDB Endowment 13(7), 2019.

[2] V. Eubanks. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press, 2018.

[3] G. Greenwald. No place to hide: Edward Snowden, the NSA, and the US surveillance state. Macmillan, 2014.

[4] S. Zuboff. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile books, 2019.

[5] V. Warmerdam. Beyond Broken - Horrible Remedies for Broken Recommenders. Published online at `https://koaning.io/posts/beyond-broken/`, 2021.

Sebastian Schelter

University of Amsterdam & Ahold Delhaize Research, The Netherlands