**EDITORIAL**

# Data Sovereignty and Data Space Ecosystems

Matthias Jarke · Boris Otto · Sudha Ram

Data analytics/artificial intelligence and business process automation require an ever-increasing wealth of data. To remain competitive, organizations cannot just use internal and publicly available data sources, but need information also from external individuals and organizations. As, e.g., supply chains evolve into highly flexible supply and demand networks, much of the required data exchange cannot be prepared any longer by lengthy human negotiations but must be semi-automatically negotiated, executed and monitored for contractual and legal compliance.

Furthermore, on top of exchanging data within a business network, an increasing number of innovative business services requires data sharing, i.e., the joint use of data from different sources within the network. An example is collaborative predictive maintenance which is based on the analysis of and learning from production process data from

Prof. Prof. (OM) Dr. M. Jarke (✉)
Information Systems, RWTH Aachen University, Ahornstr. 55, 52074 Aachen, Germany
e-mail: jarke@dbis.rwth-aachen.de

Prof. Prof. (OM) Dr. M. Jarke
Fraunhofer FIT, Aachen, Germany

Prof. Dr.-Ing. B. Otto
Industrial Information Management, TU Dortmund, Emil-Figge-Str. 91, 44227 Dortmund, Germany
e-mail: Boris.otto@isst.fraunhofer.de

Prof. Dr.-Ing. B. Otto
Fraunhofer ISST, Dortmund, Germany

Prof. S. Ram
Management Information Systems, Eller School of Business, University of Arizona, 1130 E. Helen St, Tucson, AZ 85721-0108, USA
e-mail: ram@eller.arizona.edu

several manufacturing companies using the same production machinery. Every machine operator provides data in order to benefit from better maintenance services. The more operators contribute, i.e., share their data, the greater the benefit for each one of them.

Technologies used today include, among other things, mappings between heterogeneous data source schemas, automated data transformation, fast parallel query processing and machine learning for information fusion, and blockchains for traceability. Several recent BISE special issues have addressed such topics individually.

In practice, data exchange and sharing often happen via open or semi-open platforms, around which data and service ecosystems have evolved. Such platforms can be the intended side effect of main services such as search (e.g., Google), social networking (e.g., Facebook), entertainment (e.g., Netflix), or trade (e.g., eBay, Amazon, Otto), or much more specialized regional platforms such as the smart farming ecosystem around the farm equipment provider Claas in Germany.

So far, the majority of ecosystems has been driven by keystone players. Due to network effects (the value of a network grows quadratically with the number of participants), these platforms often quickly reach monopolistic or at least oligopolistic situations.

If knowledge-intensive user organizations operate on such platforms, the enormous amount of deep knowledge in their data can create threats for the competitive advantage of whole engineering-intensive industries, and makes them susceptible to disruptive new competitors from completely different industries. However, not taking advantage of digitalization and the resulting chances for data mining and process automation is obviously not an option either.

This dilemma has caused a debate on "data sovereignty" especially in Europe where many hidden champions and large-scale market-leading enterprises, e.g., in automotive, airline, or machine-building industries reside. Data sovereignty refers to the self-determination of individuals and organizations with regard to the use of their data. In contrast to data privacy as defined, e.g., in the European General Data Protection Regulation (GDPR), which sees the citizen in a rather passive role to be protected against powers they cannot confront on an equal footing, data sovereignty aims at enabling "data richness" by clearly negotiated and strictly monitored data usage agreements. IT security methods and tools play an important socio-technical support role in enabling data sovereignty; however, they cannot replace it.

The concept of data sovereignty has spawned the idea of "alliance-driven data ecosystems" with enabling platforms. Since 2015, a worldwide alliance of companies and research organizations called International Data Space Association and a closely related series of research projects at Fraunhofer have been elaborating a standardized reference architecture for such enabling platforms, and tested it in a broad range of use cases in industrial and societal (e.g., medical or mobility) settings.

Following a detailed requirements analysis, the reference architecture includes as main components the so-called IDS Connector – a software component that annotates data to be exchanged with usage policies –, a broker, identity management, and a clearing house for data exchange and sharing transactions. This special issue includes an interview which represents this development at the highest level with Reinhold Achatz of ThyssenKrupp as the Chairman of the International Data Space Association.

The special issue contains four submitted papers addressing foundational elements that will become very important for successful alliance-driven platform ecosystems. Data exchange is addressed from an economic viewpoint as well as from the perspective of data re-purposing on the receiving side. Similar issues can occur when different concepts of business intelligence/AI algorithms and applications are co-engineered or co-used collaboratively without sufficient mutual understanding. A special case often involving cross-organizational data usage is process mining which usually involves personal data and therefore requires special attention with respect to the GDPR.

The first paper "Data Portability on the Internet – An Economic Analysis" by Wohlfarth considers the data portability aspect of data sovereignty – the GDPR-guaranteed right to move your own data from one online service to another one. This new right to escape a lock-in situation is shown to be of limited advantage to consumers and to have significant implications for data markets, as it also facilitates new entries into those markets.

The second paper "Discovering Data Quality Problems – The Case of Repurposed Data" by Zhang, Indulska, and Sadiq starts from the observation that data analytics and machine learning, especially in case of acquired external data, pursue purposes that are quite different from the original goals of the data collectors. This requires a second look at data quality from the viewpoint of the new purpose. The LANG approach in this paper employs a Design Science method on the basis of semiotic theory and data quality dimensions to develop and validate a solution of this new data quality problem.

The third paper „Privacy-Preserving Process Mining – Differential Privacy for Event Logs" by Mannhardt, Koschmider, Baracaldo, Weidlich, and Michael presents an approach to the secondary use of personal information residing in event logs. The proposed protection model uses differential privacy for process discovery methods and, thus, paves the way to leveraging a rich source of data (event logs) while preserving protection interests with regard to personal information.

Conceptual modeling research about the kind of data exchange envisioned by alliance-driven data ecosystems is still at a relatively early stage. The fourth paper "The New Area of Business Intelligence Applications – Building From a Collaborative Point of View" by Teruel, Maté, Navarro, González, and Trujillo presents a new modeling language which aims at modeling and eliciting the goals and information needs of collaborative BI systems, thus addressing one important aspect of this modeling challenge.

This issue comprises papers submitted specifically to the special issue, as well as topically closely related general submissions to BISE. The guest editors would like to express their gratitude to all reviewers, authors, and the chief editor for their important contributions to this special issue. We hope you enjoy the result and become as enthusiastic as we are about data sovereignty which we consider a critical success factor for future AI applications in knowledge-intensive inter-organizational domains.