Summary

The essential role of data in today's industrial value creation is undoubted. Data are no longer an offshoot of a physical product, but provide the basis for fundamentally new digital services and business models. Increasingly, organizations recognize that to leverage the potential of data, their cross-organizational sharing, use, and processing is necessary in interorganizational, global, and open data ecosystems. Data ecosystems promise significant benefits for businesses and society by enabling previously unimaginable data-driven services. Still, many relevant data sets are sensitive and highly protected. The development of services and approaches for realizing data ecosystems in the face of these trust issues on data sharing has grown increasingly important, in particular with data spaces that foster networks of data resources and aim to keep data holders in control of their shared data. However, an overarching ecosystem perspective reveals a lack of shared governance and interoperability among the diverse technological approaches in the growing landscape of emerging ecosystems. A distinct infrastructure for data that enables heterogeneous and distributed sharing of data without relying on closed systems or platforms is lacking.

In recent decades, market dynamics and technological solutions have emerged that emphasize the profit of individual players and pay little attention to the creation of a wide-reaching and fair playing field. From a research perspective, a set of unresolved issues and dynamics have grown: Ecosystems are prevalently addressed from a platform perspective, with only a few platform owners or complementors realizing benefits. Further, the facets and challenges of dealing with data as a stand-alone asset represents a relatively young field of research and has not yet been considered from an infrastructure perspective. While it is widely acknowledged that generic digital infrastructures cannot be designed top down or managed as a closed system, there is a great practical and even business political need for dedicated data infrastructure. Active design is urgent and necessary, and retrospective considerations alone are no longer sufficient in research. Infrastructures are also defined relationally to their usage and are not statically reducible to their components, which makes their design even more challenging.

Additionally, the networked economy reveals the need to investigate the dynamics of complex and even unpredictable systems—and how to influence and shape them. Complexity is a factor that inherently characterizes the networked world and implies many technical and managerial challenges. Complexity is often viewed as a disruptive factor that needs to be reduced, but it is unavoidable in ecosystem and infrastructure contexts.

Moreover, the associated dynamics, self-organization, evolution, and even chaotic states are essential for achieving innovation and novelty, and with that, enabling new design decisions and sustained progress. Thus, this study addresses the research question of how data infrastructures should be designed for data ecosystems. To answer this question and to do justice to the dynamics involved in designing data infrastructures, this study adopts a design science research perspective and aims to develop a set of design principles that stem from complexity science, and in particular, chaos theory. This study conceptualizes data infrastructures as complex and chaotic systems that are deterministic and sensitive to initial conditions, yet unpredictable and nonlinear while being subject to different impacts.

The study is based on a single case study of an extreme and revelatory case, the Gaia-X initiative, between 2019 and 2023. The goal is to explore the design process and results and to draw conclusions and recommendations for future design endeavors, as well as to explore how complexity and states on the edge of chaos are handled. More concretely, the results aim to provide theoretically grounded design knowledge to guide further initiatives, decisionmakers, community managers, and architects, as well as policymakers and funding bodies, so that they can make informed decisions that promote the emergence of an overarching data infrastructure. The interpretive case analysis uses an engaged scholarship paradigm and utilizes qualitative content analysis techniques—in particular, qualitative coding techniques—to identify design events. Therefore, a conceptual research framework based on chaos theory is employed as a theoretical lens to guide the investigation. The final result is a set of 16 design principles that build on empirical insights, the core characteristics of chaotic systems, and Aristotelian logic to enable a constructive design process and a sustainable design product of data infrastructures. The final design principles are based on various intermediate results, which cover an exploration of white spots in infrastructure research for data ecosystems, an empirically derived problem space from the extreme case, and a set of design areas ("key design focus topics"). The key decision points are extracted and a set of knowledge nuggets identified, which serve as a form of meta-requirements for the design principles and provide insight into the Gaia-X case. The resulting design knowledge is demonstrated by mapping to the Gaia-X 4 KI project, and represents a cross-domain project that aims to realize data flows based on the Gaia-X data infrastructure. In addition, the results and intermediate results were evaluated using semistructured interviews with 14 experts. The findings and insights of this study provide guidelines for the future design of data infrastructures, particularly in large-scale consortia.