

---

Increasingly, industry organizations are recognizing that to leverage data's potential, cross-organizational data sharing, use, and processing are necessary in data ecosystems. To enable data ecosystems, data space technologies that address relevant functionalities are emerging; however, the emerging landscape reveals a lack of shared governance, trust, and interoperability among the diverse technological approaches adopted and actors involved. Thus far, ecosystems have been prevalently addressed from a platform perspective, where few owners or complementors realize significant benefits, meaning a distinct data infrastructure for fair and balanced data ecosystems is lacking. Although the concepts of data infrastructures and dependencies in data ecosystems are becoming increasingly important for political and economic reasons, the fundamentals and targeted design of data infrastructures have not yet been investigated. Data ecosystems are inherently complex, meaning their data infrastructure must be designed for intricacy, necessitating diverse technologies for immense scalability and dynamic connections across many actors. Complexity is often viewed as a factor that must be reduced, but it is unavoidable in ecosystems, and many of its implications, including self-organization, evolution, and chaotic states, are essential to achieving large-scale innovation and novelty. Thus, this study addresses the following research question: how to design data infrastructures for data ecosystems?

To answer the research question, this thesis adopts a design science research perspective and suggests a set of design principles that stem from complexity science, particularly chaos theory. In addition, the design principles are based on a single case study of an extreme and revelatory case, namely, the Gaia-X initiative between 2019 and 2023. The interpretive case analysis follows the engaged scholarship paradigm and utilizes qualitative content analysis techniques, to identify and analyze design events. The result is a set of 16 design principles that builds on empirical insights, the characteristics of chaotic systems, and Aristotelian logic. The findings and insights provide guidelines for the future design of data infrastructures, particularly in large-scale consortia, and decision makers, including community managers and architects; and policy funding bodies so they can make informed decisions that promote the emergence of data ecosystems.