

Chapter 1

Real-time Linked Dataspaces: A Data Platform for Intelligent Systems Within Internet of Things-Based Smart Environments



Keywords Data platform · Intelligent systems · Internet of things · Data ecosystem · Smart environment · Dataspace

1.1 Introduction

Around 18,000 BCE, Paleolithic tribespeople marked notches into sticks, or bones, to keep track of trading activity or supplies. The tribespeople would compare the notches on their prehistoric data storage devices (their tally sticks) to make basic calculations that would allow them to make predictions such as how long their food supplies would last. From these early examples, we can trace a gradual evolution of the ability of humans to store, analyse, and share information.

In the first decades of the twenty-first century, datafication is driving the transformation of our everyday world, from the digitisation of traditional infrastructure (smart energy, water, and mobility) to the revolution of industrial sectors (cyber-physical systems, autonomous vehicles, and Industry 4.0), and changes to how our society works (smart government and cities). The contemporary wave of datafication is creating smart environments that are powered by digital technologies such as the Internet of things, big data, and artificial intelligence. Within these smart environments, intelligent systems are creating data ecosystems with unprecedented levels of real-time data about our world [1]. A recognition has emerged among researchers and practitioners that a new class of information management and processing systems is needed to support diverse distributed real-time data-intensive intelligent systems. These applications necessitate a transformation in how data is managed and shared among systems [2] and in how data can be processed on-the-fly and with low-latency [3]. Both of these requirements are critical if we are to extract the maximum value from the current wave of datafication, and both topics have a rich body of ongoing work. However, there is a paucity of research on approaches that tackle both of these requirements together for the large-scale sharing of real-time data.

Real-time Linked Dataspaces (RLD) address this need by combining pay-as-you-go data management with techniques for flexible data integration and real-time

processing and query. This book establishes the foundations and principles of Real-time Linked Dataspaces [4] as a data platform for intelligent systems within smart environments. It investigates the “best-effort” approximate techniques needed to process real-time data within the dataspace paradigm. The book details state-of-the-art techniques from artificial intelligence, knowledge graphs, Internet of things, and advanced stream and event processing to complement the dataspace approach to effectively and efficiently manage and extract value from data within Internet of things-based smart environments.

The remainder of this chapter is structured as follows: Sect. 1.2 begins by establishing the foundations of Real-time Linked Dataspaces with an overview of intelligent systems, smart environments, Internet of things, data ecosystems, and the need for a data platform. Section 1.3 introduces the notion of a Real-time Linked Dataspace and its role as a data platform for intelligent systems within smart environments. An overview of the structure of the book is provided in Sect. 1.4, with a summary in Sect. 1.5.

1.2 Foundations

As illustrated in Fig. 1.1, Real-time Linked Dataspaces lie at the intersection of the fields of Data Management (data ecosystems, pay-as-you-go, knowledge graphs), Distributed Systems (Internet of things, event and stream processing), Artificial Intelligence (intelligent systems), and Ubiquitous Computing (smart environments). These fields of computer science need to be brought together to enable breakthroughs not possible when the fields work in isolation. In this section, we examine the recent developments in these fields with the rise in importance of data-intensive techniques and the need to support data sharing between the ecosystems of interconnected intelligent systems within Internet of Things (IoT)-based smart environments.

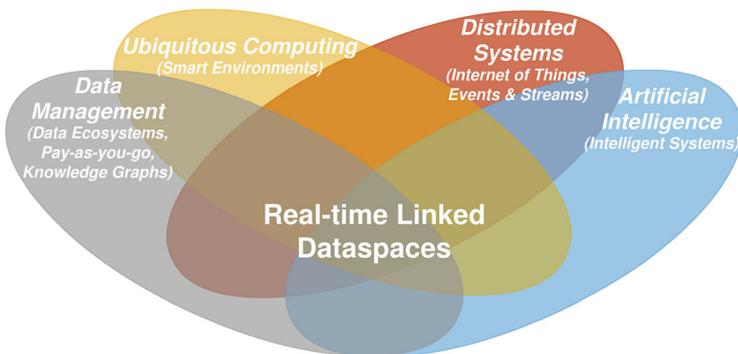


Fig. 1.1 Foundations of Real-time Linked Dataspaces

1.2.1 *Intelligent Systems*

Originating from the field of Artificial Intelligence (AI), intelligent systems are revolutionising many industries and society, including transportation and logistics, security, manufacturing, energy, healthcare, and agriculture, by providing the “built-in” intelligence to improve efficiency, quality, and flexibility. An intelligent system can gather, represent, reason, and interpret data. In doing so, it can learn, extract patterns and meaning, derive new information, learn from experience, and identify strategies and behaviours to act intelligently. Contemporary intelligent systems are usually Internet-connected with an ability to communicate and collaborate with other systems. Several definitions for intelligent systems exist with some of these captured in Table 1.1.

Intelligent systems are complex and can be created using a wide range of techniques from AI, machine learning (supervised, unsupervised, and reinforcement learning), deep learning, computer vision, natural language processing, to complex event processing, and knowledge graphs. The inspiration for the design of intelligent systems is often drawn from ideas and concepts from nature’s problem-solving approaches across a range of fields, including biology, cognitive science, and neuroscience which results in many interdisciplinary relationships. The design and construction of intelligent systems is a vibrant area of active research, which is the subject of many excellent books. The focus of this book is to support the sharing of data between intelligent systems within an IoT-enabled smart environment.

1.2.2 *Smart Environments*

Smart environments have evolved from the fields of ubiquitous and pervasive computing that promote the idea of an information communication technology-

Table 1.1 Definitions of an Intelligent System

Definition	References
“An intelligent system is one that uses artificial intelligence (AI) techniques to offer important services (e.g., as a component of a larger system) to allow integrated systems to perceive, reason, learn, and act intelligently in the real world.”	ACM Transactions on Intelligent Systems and Technology
“Intelligent systems perform search and optimization along with learning capabilities.”	[5]
“Intelligent systems connect users to artificial intelligence (machine learning) to achieve meaningful objectives. An intelligent system is one in which the intelligence evolves and improves over time, particularly when the intelligence improves by watching how users interact with the system.”	[6]
“A system, which is based on approach(es), method(s) or technique (s) of the artificial intelligence field to perform more accurate and effective operations for solving the related problems.”	[7]

Table 1.2 Definitions of a Smart Environment

Definition	References
“On a physical world richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives and connected through a continuous network.”	[8]
“An ecosystem of interacting objects, e.g. sensors, devices, appliances and embedded systems in general, that have the capability to self-organize, to provide services and manipulate/publish complex data.”	[9]
“A physical world interwoven with invisible sensors, actuators, displays, and computational elements. These computing elements are generally embedded seamlessly in everyday objects and networked to each other and beyond (the internet, usually).”	[10]
“One that is able to acquire and apply knowledge about the environment and its inhabitants in order to improve their experience in that environment.”	[11, 12]

enabled physical world. Mark Weiser defined a smart environment as “a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network” [8]. Several definitions for smart environments are detailed in Table 1.2. These definitions illustrate the complexity of the challenge associated with delivering a smart environment. Their realisation needs contributions from several research fields to be brought together to deliver on this vision, including distributed computing, mobile computing, location computing, context-aware computing, wireless sensor networks, human–computer interaction, ambient intelligence, and artificial intelligence.

In the past decade, smart environments have started to move from a research vision to concrete manifestations in real-world deployments. As smart environments are realised, they encounter a number of practical challenges, including the interoperability of diverse technology (e.g. legacy systems) [13], meeting the needs of diverse stakeholders with very broad goals and expectations, and working within the limited budgets available to invest in infrastructure. The understanding of these challenges in more detail requires an understanding of how the IoT is enabling real-time data processing within smart environments.

1.2.3 *Internet of Things*

A key driver in the development of smart environments is the convergence of technologies such as the IoT and big data, which is driving the digitisation of physical infrastructures with sensors, networks, and social capabilities [14]. The vision of the IoT is a complicated proposition that requires end-to-end distributed systems from the development of new electronic devices and embedded systems, new forms of data processing to deal with the volume, variety, and velocity of data

generated, to enhanced user experiences leveraging cognitive and behavioural models with new data visualisation and interaction paradigms.

As the IoT enables the deployment of lower-cost sensors, we see more broad adoption of IoT devices/sensors and gain more visibility (and data) into smart environments. This results in high volume and high-velocity event streams from smart environments that need to be processed. IoT-based smart environments are also generating different types of data with an increase in the number of multimedia devices deployed, such as vehicle and traffic cameras. The IoT is driving the deployment of intelligent systems and creating new opportunity in smart environments:

- **Digital Twins:** A digital replica of physical assets (car), processes (value-chain), systems, or physical environments (building). The digital representation (i.e. simulation modelling or data-driven model) provided by the digital twin can be analysed to optimise the operation of the “physical twin”.
- **Physical-Cyber-Social (PCS):** A computing paradigm that supports a richer human experience with a holistic data-rich view of the smart environment that integrates, correlates, interprets, and provides contextually relevant abstractions to humans [14].
- **Mass Personalisation:** More human-centric thinking in the design of systems where users have growing expectations for highly personalised digital services for the “Market of One”.
- **Data Network Effects:** As more systems/users join and contribute data to the smart environment, a “network effect” can take place, resulting in the overall data available becoming more valuable.

Within this context, we are interested in how data created within a smart environment can be leveraged by intelligent systems, and how data can be easily shared within the ecosystem of systems (new and old) and stakeholders.

1.2.4 Data Ecosystems

A Data Ecosystem is a socio-technical system enabling value to be extracted from data value chains supported by interacting organisations and individuals [15]. Within an ecosystem, data value chains are oriented to business and societal purposes. The ecosystem can create the conditions for a marketplace competition among participants or enable collaboration among diverse, interconnected participants that depend on each other for their mutual benefit.

The digital transformation is creating a data ecosystem with data on every aspect of our world, spread across a range of intelligent systems. As illustrated in Fig. 1.2, a smart environment enabled with IoT data, and contextual data sources, results in a data-rich ecosystem of structured and unstructured data (e.g. images, video, audio, and text) that can be exploited by data-driven intelligent systems.

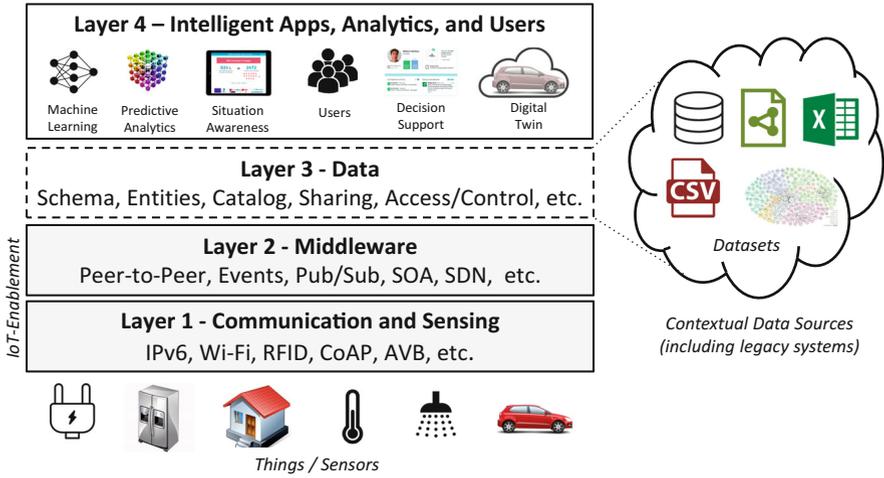


Fig. 1.2 Four-layered framework to enable data ecosystems for intelligent systems within IoT-based smart environments [4]

There is a need to bring together data from the multiple intelligent systems that exist within the data ecosystem surrounding a smart environment. For example, smart cities are showing how different systems within the city (e.g. energy and transport) can collaborate to maximise the potential to optimise overall city operations. At the level of an individual, digital services can deliver a personalised and seamless user experience by bringing together relevant user data from multiple systems [16]. This requires a System of Systems (SoS) approach to connect systems that cross organisational boundaries, come from various domains (e.g. finance, manufacturing, facilities, IT, water, traffic, and waste) and operate at different levels (e.g. region, district, neighbourhood, building, business function, individual).

Data ecosystems present new challenges to the design of intelligent systems and SoS that require a rethink in how we should deal with the needs of large-scale, data-rich smart environments. How can we support data sharing between intelligent systems in a data ecosystem? What are the technical and non-technical barriers to data sharing within the ecosystem? How can intelligent systems leverage their data ecosystem to be “smarter”? Solving these problems is critical if we are to maximise the potential of data-intensive intelligent systems [1].

1.2.5 Enabling Data Ecosystem for Intelligent Systems

Understanding the data management challenges in more detail requires an appreciation of how the IoT is enabling smart environments. The range of IoT challenges can be studied based on the three-layered framework by Atzori et al. [17] *Layer 1—Communication and Sensing; Layer 2—Middleware; and Layer 3—Users, Applications, & Analytics*. At the data level, intelligent systems can benefit from leveraging

data from multiple systems within the smart environment. However, many of the data management and sharing activities are currently performed at the application layer within IoT deployments. We elaborate a four-layered framework to enable data ecosystems for intelligent systems within IoT-based smart environments that builds on the work by Atzori et al. [17]. As illustrated in Fig. 1.2, we introduce a fourth layer between the Middleware and Application layers to support data management and sharing activities. The four-layered framework for enabling data ecosystems for intelligent systems consists of:

- **Layer 1—Communication and Sensing:** An essential requirement is an infrastructure of communication and sensing that maps the world of physical things into the world of computationally processable data.
- **Layer 2—Middleware:** Middleware abstracts the application developers from the underlying technologies. Data distribution, processing, and access to legacy information systems take place at this layer.
- **Layer 3—Data:** There is a need to enable data management and sharing activities, including managing schema and entities, accessibility, access control, data quality, and licensing take place at this layer.
- **Layer 4—Intelligent Applications, Analytics, and Users:** Users expect IoT-based analytics and applications that present the data gathered and analysed in an intuitive and user-friendly manner using new visualisations and user experiences to ensure cognitive-friendly smart environments.

Our key addition is *Layer 3—Data*, which requires the development of data infrastructure to support the sharing and management of data among systems in the ecosystem. Platform approaches have proved successful in many areas of technology, and the idea of large-scale “data” platforms are touted as a possible next step. A data platform focuses on secure and trusted data sharing amongst a group of participants (e.g. industrial consortiums sharing private or commercially sensitive data) within a clear legal framework. Within a smart environment, a data platform would support continuous, coordinated data flows, seamlessly moving data among intelligent systems.

1.3 Real-time Linked Dataspaces

In this book, we advocate the use of the dataspace paradigm to support the sharing of data between intelligent systems within IoT-enabled smart environments. The dataspace approach recognises that in large-scale integration scenarios, involving thousands of data sources, it is difficult and expensive to obtain an upfront unifying schema across all sources [2]. Dataspaces are not a data integration approach [19]; they shift the emphasis to providing support for the co-existence of heterogeneous data that does not require a significant upfront investment into a unifying schema. Data is integrated on an “as-needed” basis with the labour-intensive aspects of data integration postponed until they are required. Dataspaces reduce the initial effort required to set up data integration by relying on automatic matching and mapping

generation techniques. This results in a loosely integrated set of data sources. When tighter semantic integration is required, it can be achieved in an incremental “pay-as-you-go” fashion by detailed mappings among the required data sources.

The Real-time Linked Dataspace (RLD) is a platform for data management for intelligent systems within smart environments that combines the pay-as-you-go paradigm of dataspace, linked data, and knowledge graphs with entity-centric real-time query capabilities [4]. In order to enable the dataspace principles to support real-time data processing, we created a specialised dataspace support service for loose administrative proximity and semantic integration for event and stream systems. This requirement forms the foundation of the techniques and models used to process events and streams within RLD.

The RLD contains all the relevant information within a data ecosystem including things, sensors, and data sources and has the responsibility for managing the relationships among these participants. The RLD goes beyond a traditional dataspace approach by supporting the management of entities within the data ecosystem as first-class citizens along with data sources, and it extends the dataspace support platform with real-time processing and querying capabilities. Figure 1.3 illustrates the architecture of the RLD with the following central concepts:

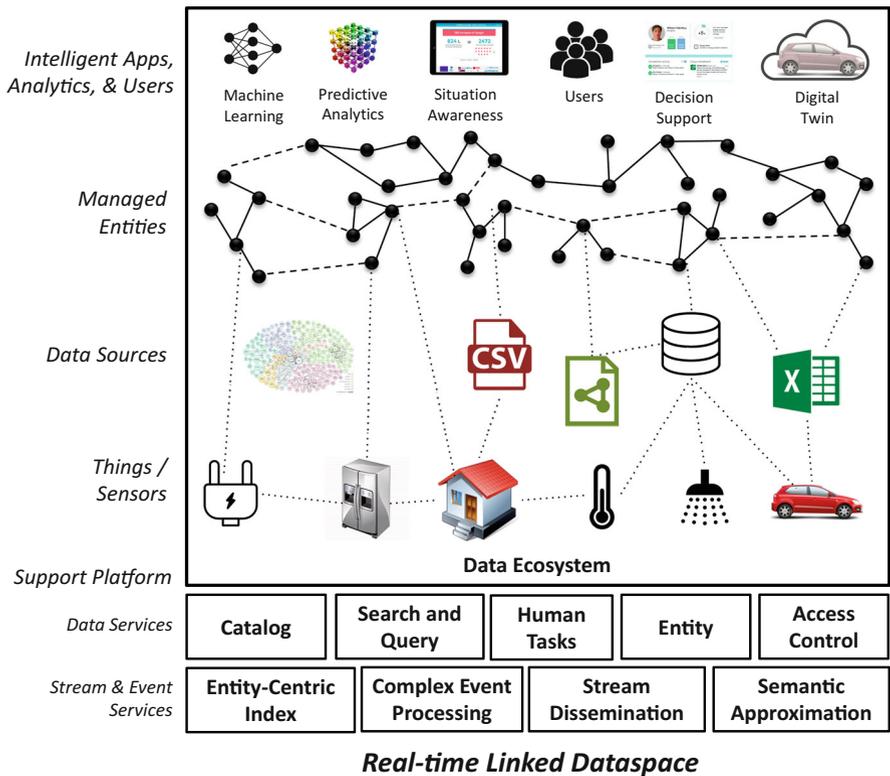


Fig. 1.3 Real-time Linked Dataspace architecture [4]

- **Support Platform:** Responsible for providing the functionalities and services essential for managing the dataspace. Support services are grouped into data services and stream and event services.
- **Things/Sensors:** Produce real-time data streams that need to be processed and managed. Things in a smart environment range from connected personal devices and sensors to connected cars and manufacturing equipment.
- **Data Sources:** Data can be available in a wide variety of formats and accessible through different system interfaces. Some examples of data sources include building management systems, energy and water management systems, personal information systems, enterprise databases, weather forecasts, and (linked) open data.
- **Managed Entities:** Actively managed entities within the data ecosystem, including their relationship to participating things, data sources, and other entities in the RLD.
- **Intelligent Applications, Analytics, and Users:** Interact with the RLD and leverage its data and services to provide data analytics, decision support tools, user interfaces, and data visualisations. Applications/users can query the RLD in an entity-centric manner, while users can be enlisted in the curation of the data and entities via the Human Task service.

The RLD has been used as a data platform to support the development of intelligent applications within a range of IoT-based smart environments including smart home, school, office building, university, and airport [16]. Within these environments, a data platform needs to support a wide range of end-users with different interests and priorities; from corporate managers looking for data to improve the performance of their business to software engineers developing intelligent applications for smart environments (see Fig. 1.4).

1.4 Book Overview

This book brings together the body of work on Real-time Linked Dataspaces and structures it (as illustrated in Fig. 1.5) into four parts:

- The first part of the book details the motivation and core concepts of Real-time Linked Dataspaces. This part explores the need for an evolution of data management techniques to meet the challenges of data ecosystems for intelligent systems. Chapters in part I cover knowledge sharing among intelligent systems in data ecosystems, fundamentals of the dataspace approach to data management, and introduce the Real-time Linked Dataspace concept and its role as a data platform for intelligent systems within IoT-enabled smart environments.

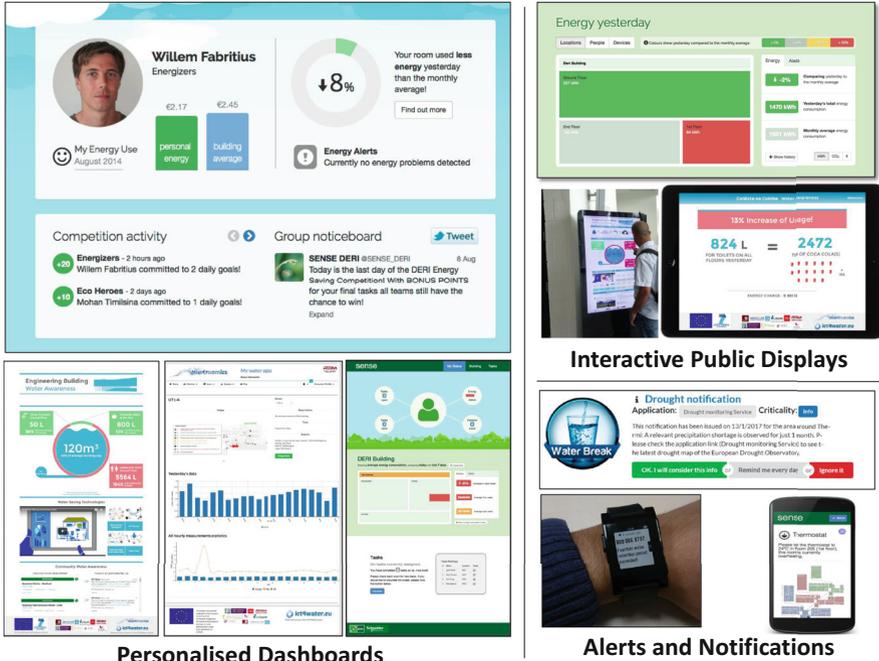


Fig. 1.4 Intelligent applications built using the Real-time Linked Dataspace [16]

- The second part of the book explores the essential data management support services provided by the Real-time Linked Dataspace. Part II contains chapters that detail data services, including catalog, entity management, query and search, data discovery, and human tasks.
- The third part of the book explores advanced stream and event processing support services for Real-time Linked Dataspaces. Chapters detail advanced techniques for approximate and best-effort stream and event processing services for dataspaces including quality of service, complex event processing, dissemination of IoT streams, and approximate semantic event matching.
- The fourth part of the book explores the use of Real-time Linked Dataspaces within real-world smart environments. The chapters in this part demonstrate the role of the Real-time Linked Dataspace in enabling intelligent water and energy management systems through the development of IoT-based digital twins and intelligent applications, IoT-enhanced user experience, and autonomic source selection for advanced predictive analytics.
- The final part of the book discusses what is required for the widespread adoption of the Real-time Linked Dataspace approach and details a future research agenda for dataspaces, data ecosystems, and intelligent systems.

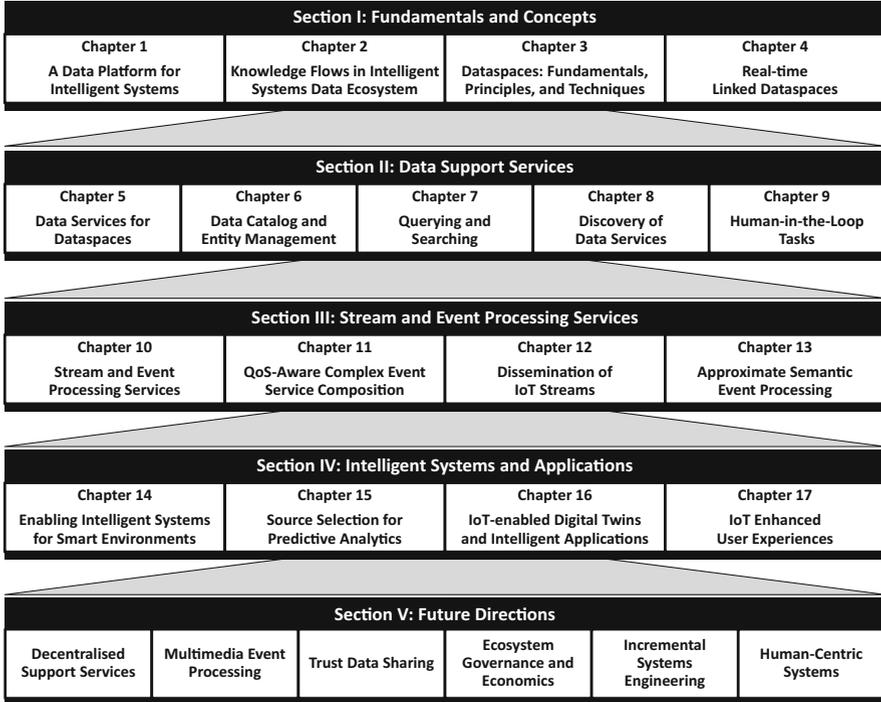


Fig. 1.5 Structure of the book

1.5 Summary

In this chapter, we have established the growing importance of intelligent systems as our society undergoes a digital transformation. Internet of Things enabled smart environments will generate vast amounts of data that create new opportunities for intelligent systems. This book postulates the use of the dataspace data management paradigm as the core of a data platform to enable data ecosystems for intelligent systems.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

