5.1 Introduction

The objective of a Real-time Linked Dataspace is to support a real-time response from intelligent systems to situations of interest when a set of events take place within a smart environment. In addition to the obvious need for real-time data processing support services, there is also the need for the fundamental data support services one would expect in a dataspace support platform. This part of the book discusses the enhanced data support services developed for the Real-time Linked Dataspace to support data management for intelligent systems within smart environments. The goal of these services is to support a real-time dataspace system to get up and running with a low overhead for administrative setup costs (e.g. catalog, entity management, search and query, and data service discovery). Each of the support services has been specifically designed for and evaluated within Internet of Things-based smart environments. This chapter provides a high-level overview of the data support services discussed in Part II and details their tiered service levels. This chapter is structured as follows: Sect. 5.2 provides a brief overview of the pay-as-you-go data support services covered in this part of the book, while Sect. 5.3 details how the services support the 5 star scheme. A summary is provided in Sect. 5.4.

5.2 Pay-As-You-Go Data Support Services for Real-time Linked Dataspaces

The adoption of the Internet of Things (IoT)-enabled smart environments is empowering data-driven systems that are transforming our everyday world. To support the interconnection of intelligent systems in the data ecosystem that
surrounds a smart environment, there is a need to enable the sharing of data among intelligent systems. A data platform can provide a clear framework to support the sharing of data among a group of intelligent systems within a smart environment [1] (see Chap. 2). In this book, we advocate the use of the dataspace paradigm within the design of data platforms to enable data ecosystems for intelligent systems.

We have created the Real-time Linked Dataspace (RLD) (see Chap. 4) as a data platform for intelligent systems within smart environments. The RLD combines the pay-as-you-go paradigm of dataspaces with linked data, knowledge graphs, and real-time stream and event processing capabilities to support large-scale distributed heterogeneous collection of streams, events, and data sources [4]. At the foundation of the pay-as-you-go approach to data integration is the idea that the owners of the data sources are responsible for the incremental improvement in the integration and quality of data available in the dataspace. The needs of the user drive incremental improvements over time. This pragmatic approach allows the dataspace to grow and enhance gradually with data sources or streams joining or leaving at any time. In order to reduce the burden on data source owners and users of the RLD, a support platform with a number of data support services is provided.

The design of the support services needs to conform to the principles of RLDs. The RLD principles specialise the dataspace principles as set out by Halevy et al. [78] to describe the specific requirements within a real-time dataspace setting:

- A Real-time Linked Dataspace must deal with many different formats of streams and events.
- A Real-time Linked Dataspace does not subsume the stream and event processing engines; they still provide individual access via their native interfaces.
- Queries in the Real-time Linked Dataspace are provided on a best-effort and approximate basis.
- The Real-time Linked Dataspace must provide pathways to improve the integration among the data sources, including streams and events, in a pay-as-you-go fashion.

The data services provided by RLD (Fig. 5.1) are:

![Fig. 5.1 Data services provided by the real-time linked dataspace](image)
• **Catalog:** The catalog service plays a crucial role by providing information about participating data sources in the dataspace. Within the catalog, all datasets and entities are declared along with relevant metadata.

• **Entity Management:** The Entity Management Service (EMS) manages information about the entities (e.g. real-world objects) in the dataspace. The EMS is an essential service for decision-making applications that rely on accurate entity information.

• **Search and Query:** The Search and Query services help developers, data scientists, and users to find relevant data sources within the dataspace.

• **Data Service Discovery:** Efficiently describing and organising data sources in dataspaces is essential. The Data Service Discovery Service organises and indexes data sources based on their capabilities.

• **Human Tasks:** The Human Task service is concerned with the collaborative aspect of data management within the dataspace by enabling small data management tasks (e.g. data quality and enrichment) to be distributed among users in the smart environment. The Human Task service can also engage participants in citizen actuation tasks within the smart environment.

An essential requirement for intelligent systems within a smart environment is to support the querying of real-time data streams. Within the RLD this is achieved by several support services for processing streams and events which are covered in Part III of this book. In the remainder of this part of the book, we detail the above data support services and focus on how they enable data management in the RLD. Each of these services has been designed to follow the RLD principles and to offer tiered service-levels following the 5 star pay-as-you-go model from Chap. 3.

### 5.3 5 Star Pay-As-You-Go Levels for Data Services

The 5 star scheme details the level of integration of the data sources with the support services of a dataspace. At the 1 star level, a data source needs to be made available with the dataspace. Over time, the level of integration with the support services can be improved in an incremental manner on an as-needed basis. The more the investment is made to integrate with the support services, the better is the integration achievable in the dataspace. The different service tiers of the RLD data support services are detailed in Table 5.1.

### 5.4 Summary

This chapter provides an overview of the enhanced data support services developed for the Real-time Linked Dataspace to enable intelligent systems within IoT-based smart environments. The goal of these services is to support Real-time Linked
Table 5.1 The 5 star pay-as-you-go model for the Real-time Linked Dataspaces Support Services [4]

| Pay-as-you-go star rating | Data format | Catalog | Access control | Search and query | Entity | Human tasks |
|---------------------------|-------------|---------|----------------|-----------------|-------|-------------|
| * Basic                   | Any format (e.g. PDF). | Registry of datasets, and streams | None | Browsing | None | None |
| ** Machine-readable      | Machine-readable structured data (e.g. Excel) Documentation to understand the data/stream structure, format, and characteristics. | Non-machine-readable metadata document (e.g. PDF) | Coarse-grained (Dataset level) | Keyword search | Entities identifiers in documentation | Schema mapping |
| *** Basic integration    | Non-proprietary format (e.g. CSV, JSON, XML) | Machine-readable metadata Equivalence among schema concepts | Fine-grained (Entity-level) Secure query service | Structure search | Source level (siloed) | Entity mapping |
| **** Advanced integration| Open standards (RDF, JSON-LD) to identify things/entities using the first two principles of linked data | Relations among schemas (dataspace level) | Data anonymisation | Structured queries | Canonical identifiers and entity mappings across sources | Entity enrichment |
| ***** Full semantic integration, search, and query | Follows all publishing principles of linked data | Full semantic mappings | Usage control | Schema-agnostic question answering | Knowledge graphs semantically link entities to related entities, data, and streams | Data quality improvement |
Dataspaces to get up and running within a smart environment with a low overhead for administrative setup costs (e.g. catalog, entity management, search and query, and data service discovery). The services follow the 5 star pay-as-you-go model for tiered services.

**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.