Smart City Urban Heat Monitoring using a Solid-based Dataspace

Florian Hölken[1], Alexander Paulus[1], Tobias Meisen[1] and André Pomp[1]

[1]Institute for Technologies and Management of Digital Transformation, University of Wuppertal, Wuppertal, Germany

Abstract

The exponential growth of urban environmental data, originating from IoT sensors and weather stations, presents both opportunities and challenges, particularly in cities, where addressing climate-related issues requires the integration of heterogeneous datasets. This paper explores the use of a Solid dataspace to enhance data interoperability, sovereignty, and integration for urban heat island (UHI) monitoring. For that, we introduce a Heat Monitoring App that leverages a Solid-based dataspace to integrate and provide decentralized temperature data from various possible stakeholders, such as government agencies, environmental authorities, and citizen-generated sources. Our approach employs semantic modeling on each data source to address interoperability challenges, enabling effective data discovery and integration while preserving data sovereignty. Through an experimental evaluation, we demonstrate the feasibility of this approach in a simulated urban setting. The findings highlight the potential of Solid-based dataspace for smart city applications, while also identifying limitations related to real-time data processing, automation in schema alignment, and privacy-preserving access control. Future work aims to enhance automation, scalability, and real-time capabilities, ultimately bridging the gap between theoretical advancements in dataspaces and their practical implementation in urban climate resilience.

Keywords

Solid, Dataspaces, Urban Heat Islands, Semantic Web, Data Interoperability

This paper addresses (Issue #11) of the W3C Dataspaces Community Group.

1. Introduction

The volume of urban environmental data is growing exponentially, driven by the proliferation of IoT devices and digitalization of public services. For instance, the number of connected sensors is projected to exceed 30 billion by 2030 [1], creating vast opportunities for innovation in areas such as climate resilience. However, much of this data remains underutilized due to a lack of structured and interoperable infrastructures. Without properly designed data infrastructures, data remains trapped in isolated silos, or its utilization requires extensive effort, limiting its accessibility and impact.

To address this challenge, dataspaces have gained increasing attention in recent years, fostering data-driven innovation across various domains [2]. These structured data ecosystems enable controlled data sharing and interoperability, promoting efficient data exchange between stakeholders. Dataspaces have emerged in multiple sectors, including industry, energy and utilities, agriculture as well as smart cities. Enterprise solutions such as Catena-X, International Data Spaces (IDS) or Gaia-X provide frameworks for secure and scalable data sharing within these sectors. Solid-based dataspaces leverage the Solid technology, a web standard for decentralized data storage and access control, allowing users to store their personal data in so-called Solid Pods. These Pods are user-controlled data containers that ensure data sovereignty and portability [3]. These kind of dataspaces can be utilized in both enterprise and public dataspaces, supporting various use cases that require enhanced data sovereignty and interoperability [4]. In the context of smart cities, Solid-based dataspaces are particularly promising as they enable citizen participation, a feature often lacking in enterprise-driven solutions.

The Third International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, June 01, 2025, Portorož, Slovenia

hoelken@uni-wuppertal.de (F. Hölken); paulus@uni-wuppertal.de (A. Paulus); meisen@uni-wuppertal.de (T. Meisen); pomp@uni-wuppertal.de (A. Pomp)

0009-0002-5191-901X (F. Hölken); 0000-0002-9899-6736 (A. Paulus); 0000-0002-1969-559X (T. Meisen); 0000-0003-0111-1813 (A. Pomp)

The Solid project in general is initiated by Tim Berners-Lee and aims to reshape the current web by putting users in control of their data. It introduces a decentralized architecture for storing and managing personal and organizational data via user-controlled Pods, promoting data interoperability, privacy, and data sovereignty. By decoupling data from applications, Solid enables flexible and ethical data reuse across systems and domains.

Despite the rising trend of dataspaces and related initiatives, there is still a lack of concrete applications that leverage dataspaces as a fundamental infrastructure for developing solutions for concrete problems. While many dataspaces focus on data governance and access control, practical implementations that utilize these infrastructures for specific urban challenges remain scarce.

A critical urban challenge where heterogeneous public-sector data is crucial is the identification of urban heat islands (UHIs). The rapid urbanization of the 21st century has led to unprecedented environmental challenges, with cities becoming hotspots for climate change effects. Among these, UHIs are particularly concerning, as they exacerbate extreme heat events, disproportionately affecting urban populations and straining essential resources such as energy and water [5, 6]. The increasing frequency and intensity of heatwaves, as projected by the IPCC, underscore the urgent need for improved urban heat monitoring and mitigation strategies [7]. While digital tools for environmental monitoring exist, they often struggle with integrating heterogeneous data from multiple sources, reducing their effectiveness in providing actionable insights [8]. Addressing these challenges necessitates innovative approaches that enhance data interoperability, contextual understanding, and analytical capabilities to support informed urban planning and decision-making.

To effectively monitor UHIs, data from diverse public-sector stakeholders is required. High-resolution temperature data, alongside other environmental parameters, must be sourced from meteorological agencies, environmental authorities, research institutions, and even citizen-generated data from personal weather stations and IoT devices. However, the disparate formats and standards of these datasets pose challenges to integration and analysis.

In this paper, we demonstrate a concrete application for UHI identification by implementing an urban heat monitoring application leveraging a Solid-based dataspace. Figure 1 provides an overview of the developed application. In order to develop the application, we therefore introduce a first prototype of a decentralized Solid-based dataspace consisting of multiple solid servers. Using this dataspace our application integrates and visually presents heterogeneous temperature data from various stakeholders. By leveraging the principles of dataspaces, we enhance the interoperability and usability of urban heat data for decision-makers and citizens.

In order to enable this interoperability, a key contribution of this work is to show the role of semantics in dataspaces. For that, we examine how semantic technologies can enhance the the integration, interpretation, and usability of data in dataspaces and for the corresponding application development. By bridging the gap between raw data and actionable insights, our approach advances the discourse on smart city innovations and contributes to more effective and sustainable urban planning.

This work is conducted as part of the funded project Gesundes Tal, which aims to improve health prevention in Wuppertal by addressing environmental challenges, such as UHIs, through digitally integrated solutions. The project explores innovative approaches to environmental monitoring, leveraging Semantic Web technologies and decentralized data infrastructures to bridge the gap between theoretical advancements and practical applications in urban data ecosystems.

2. Related Work

The increasing interest in smart city solutions has led to a growing focus on data interoperability and dataspaces as a key enabling infrastructure. This section examines existing dataspaces, the role of semantic technologies, and real-world applications that leverage these technologies within dataspaces.

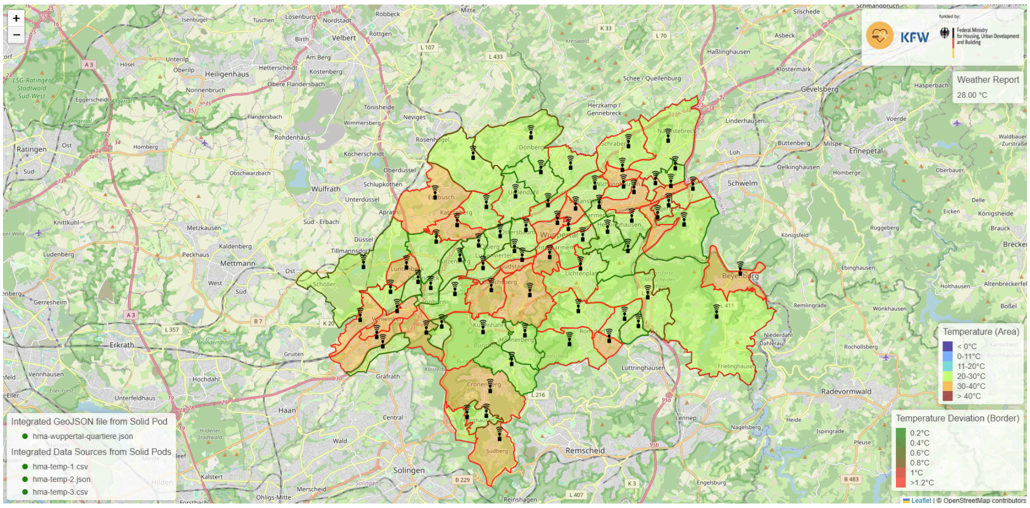

Figure 1: Showcase of the Heat Monitoring App showing possible UHIs in the city of Wuppertal.

2.1. (Solid) Dataspaces

Dataspaces have emerged as an approach for managing and integrating heterogeneous data from multiple stakeholders, particularly in domains requiring cross-organizational data exchange, such as smart cities and environmental monitoring. Originally introduced as a flexible strategy for integrating diverse data sources [9], the concept has evolved into real-world implementations such as Gaia-X and IDS, which provide frameworks for secure, interoperable data sharing across industries [10].

However, many of these existing dataspace solutions primarily cater to enterprise use cases, emphasizing industrial data governance and sector-specific interoperability [11, 10]. In contrast, smart city applications present a fundamentally different challenge: they require active citizen participation, decentralized governance and interoperable data infrastructures to effectively integrate public-sector, private-sector, and citizen-generated data. This distinction highlights the limitations of centralized dataspace architectures in smart cities and underscores the potential of Solid-based dataspaces as a more suitable alternative.

Unlike traditional dataspace architectures, a Solid-based dataspace leverages web-native (semantic) standards to enable decentralized, user-controlled data management [3]. Key standards include the Resource Description Framework (RDF), Linked Data principles, which are a set of best practices proposed by Tim Berners-Lee for publishing and interlinking structured data on the Web [12], and WebID, a decentralized identity mechanism for user authentication and access control in Solid Pods. This decentralized approach aligns closely with the requirements of smart city applications, particularly in domains such as urban heat monitoring, where data originates from multiple heterogeneous sources, including:

- Government agencies (meteorological offices, city planning departments)
- Environmental authorities (air quality and climate research institutions)
- Private organizations (utility providers, sensor manufacturers)
- Citizen-generated data (personal weather stations, IoT devices)

The ability of a Solid-based dataspace to facilitate direct citizen participation, while ensuring data sovereignty and privacy, presents a unique advantage over enterprise-focused dataspace solutions like Gaia-X or IDS. By allowing individuals and organizations to store their own environmental data in personal Solid Pods, a collaborative, yet decentralized model for urban data management can emerge.

2.2. Semantic Technologies in Dataspaces

Semantic technologies play a pivotal role in enhancing data interoperability, integration, and discoverability within dataspaces. By leveraging ontologies, semantic annotations like semantic labels or models, and Linked Data principles, these technologies enable automated and contextual understanding of heterogeneous datasets [13]. In the context of urban heat monitoring and smart city applications, semantic modeling helps us to align disparate environmental data sources, facilitating simple data model understanding and informed decision-making.

In general, semantic modeling provides expressive meaning and context to data by establishing mappings between data attributes and ontology concepts [14]. This process is crucial in heterogeneous data environments, such as smart cities, where data originates from multiple stakeholders with varying formats and standards. The semantic modeling workflow consists of two primary phases:

- **Semantic Labeling** – Establishes initial mappings between structured datasets and concepts within a predefined ontology [15].
- **Semantic Refinement** – Enhances these mappings by incorporating contextual relationships and improving alignment across diverse data sources [16, 17].

To facilitate the semantic mapping process, semi-automatic approaches have been developed. For example, Knoblock et al. [18] proposed methodologies that leverage structured data sources and integrate them into the Semantic Web, reducing manual efforts and improving the efficiency of data annotation.

Recent advances have focused on improving the creation of semantic models, making large-scale semantic integration more feasible. Systems like PLASMA [19, 16] provide platforms that combine semantic labeling with tools for manual refinement, streamlining the creation of semantic models. Similarly, scalable approaches such as those proposed by Taheriyan et al. [15, 20] demonstrate how Semantic Web principles can be used to infer semantic relationships automatically, thereby enhancing data alignment across different domains.

2.3. Real World Applications utilizing Dataspaces

While the concept of dataspaces has been widely discussed in research and industry, relatively few real-world applications demonstrate their effectiveness in solving domain-specific challenges. Most dataspace initiatives focus on data interoperability, governance, and security, but their concrete application to specific urban or environmental problems remains limited. This subsection examines existing dataspace implementations across different domains, highlighting their relevance to smart cities and the role of semantic technologies in enhancing their capabilities.

One of the most prominent real-world applications of dataspaces is Catena-X, a Gaia-X initiative aimed at creating an interoperable data infrastructure for the automotive industry [21]. This dataspace facilitates secure, cross-company data exchange, ensuring supply chain transparency and efficient resource allocation. Similarly, IDS provides a framework for standardized and secure data sharing across multiple industrial sectors [22]. However, while these initiatives demonstrate the feasibility of data interoperability, they primarily focus on enterprise and supply chain use cases, lacking applications that directly benefit end users.

In environmental monitoring, Semantic Web technologies have been explored to facilitate cross-domain data integration and improve the discoverability of environmental resources. Schimak et al. [23] demonstrated in 2013 how a semantic discovery framework, leveraging web-based data interlinking principles, can enhance access to heterogeneous environmental datasets by aligning domain-specific ontologies and enabling cross-domain search. However, no application was developed to showcase the practical implementation of this approach.

Dataspaces have been recognized as a promising concept in agriculture and healthcare, yet they remain largely unexplored in practical implementations. Wolfert et al. [24] analyzed the role of big data and semantic data models in precision farming, demonstrating how improved data sharing among farmers, agribusinesses, and policymakers could enhance decision-making and efficiency. However, while these approaches highlight the need for interoperable data ecosystems, dataspaces as an enabling infrastructure have not yet been realized in this domain nor any applications.

Despite the growing adoption of dataspaces in various domains, there is a clear gap in their application to urban heat monitoring and smart city climate resilience. Existing implementations predominantly focus on governance, industry, and large-scale data infrastructures, while solutions that empower cities and citizens to leverage decentralized, interoperable climate data remain underdeveloped. This study addresses this gap by demonstrating an urban heat monitoring application that leverages a Solid-based dataspace, combining the semantic interoperability advantages of Linked Data with the decentralized, privacy-preserving nature of Solid Pods.

By leveraging these technologies, we aim to show that dataspaces can be a viable solution for integrating and utilizing heterogeneous environmental data, ultimately contributing to more actionable climate adaptation strategies in smart cities among different participants.

3. Methodology & Implementation

As discussed above, this study aims to develop a Heat Monitoring App for identifying UHIs using a Solid-based dataspace, demonstrating a practical application within the context of dataspaces. By leveraging semantic technologies in the form of semantic models, our approach aims especially for better data discovery and understanding for application developers. It also helps the developer to transform fragmented and heterogeneous datasets into meaningful representations, supporting advanced analysis and actionable insights for urban heat monitoring.

3.1. Setting up a Solid-based Dataspace

As mentioned above, we decided to use Solid as the foundational technology for setting up our dataspace. The choice of Solid for our work is rooted in the need for a participatory model that accommodates diverse stakeholders for our real-world application. In addressing urban heat challenges, a dataspace must support collaboration among city administrations, organizations, and citizens.

Compared with the functionalities described in the International Data Spaces Association (IDSA) reference architecture, Solid offers several features directly out of the box, as highlighted by Harth et al. [3]. These include decentralized data storage through Pods, access control mechanisms, and Semantic Web compatibility for machine-readable data representation and integration.

Our envisioned Solid-based dataspace is a distributed environment where each stakeholder manages their data independently within their Solid Pods. These Pods serve as modular, secure storage units, allowing users to maintain full control over their data. Using WebID, the system ensures decentralized identity management, streamlining authentication and enabling flexible Access Control Policies (ACP) for privacy-preserving data sharing. We chose to adopt ACP because it offers finer-grained, more flexible, and extensible access control capabilities, and it is aligned with the ongoing evolution of the Solid standard.

To implement this infrastructure, we utilize the Community Solid Server (CSS) as the backend for hosting and managing Solid Pods. The CSS provides a flexible and open-source implementation of the Solid protocol, allowing for decentralized data storage and retrieval while ensuring compliance with Solid standards.

3.2. Implementation of a data catalog

However, while Solid provides many essential building blocks, one key component is missing: a data catalog for organizing and discovering datasets efficiently. In general, a data catalog within a dataspace serves as a central component for enabling advanced data management. Its primary goal is to organize and enrich metadata to facilitate data discovery, contextual understanding and interoperability across diverse datasets. Semantics play an important role in achieving this by providing a shared vocabulary and structure that allows heterogeneous data sources to be discovered and integrated. To enable this, two categories of information must be modeled: (1) metadata about datasets, and (2) the semantic content of the datasets themselves.

For the first category, we employ the DCAT (Data Catalog Vocabulary) standard to manage and describe the metadata of data sources. DCAT, developed by W3C, is widely adopted for cataloging datasets, distributions and services in data ecosystems. It provides a structured approach to metadata representation, supporting data discovery and alignment with dataspace guidelines.

For the second category, we utilize semantic models, as outlined by Knoblock et al. [25, 14], to represent the content of the data in a semantically rich and interoperable manner. While DCAT forms the backbone of the metadata structure, it is extended with semantic models to capture contextual information about the data assets. These semantic models, expressed in RDF, provide additional layers of meaning, allowing users to understand the relationships, provenance and constraints of the datasets. Figure 2 illustrates the integration of the semantic model within the data catalog.

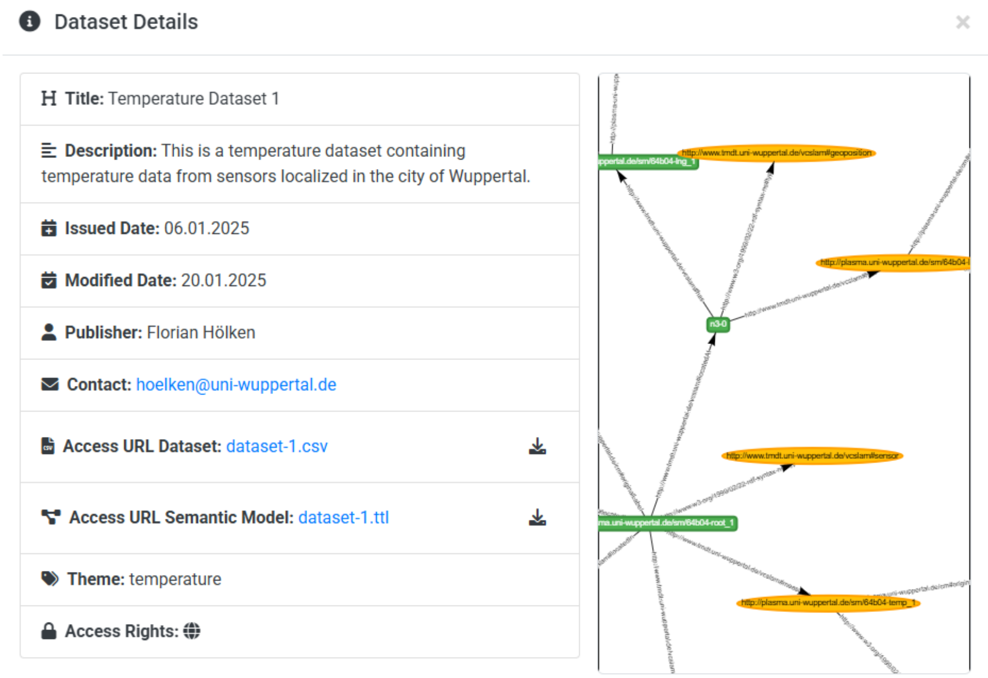

Figure 2: Implementation of the integration of the semantic model in the data catalog.

A key design decision for our Solid-based dataspace is the retention of original data formats. Unlike traditional semantic systems that convert all data to RDF, we decided for our dataspace that all data remains in its native formats, such as JSON or CSV. This decision was motivated by two factors: First to ensure practical integration with existing sensor infrastructures and public datasets, which predominantly provide non-RDF data, and second to lower the technical entry barriers for stakeholders who might not be familiar with RDF. We designed our data catalog accordingly to enable semantic enrichment and integration without requiring full data format conversion. This approach aligns with practical requirements and industry standards, such as those outlined by the IDSA, which generally do not mandate RDF-based data transformation [22]. Instead, the semantic layer focuses on representing contextual information in RDF, utilizing semantic modeling to bridge the gap between non-RDF data formats and semantic reasoning capabilities.

Although the catalog currently focuses on indexing publicly available datasets from Solid Pods, we acknowledge that exposing dataset URLs may raise minor privacy concerns, such as the potential linkability or re-identification of information. To mitigate these risks, future work will explore access-controlled catalog entries that balance discoverability with the principles of data sovereignty and protection.

3.3. Heat Monitoring App

The Heat Monitoring App is designed to function as a consumer of data from the dataspace and contributor of visually presented information to the end user. By integrating into the decentralized environment, the app developer leverages the data catalog to discover and access relevant data assets. The use of semantic models enables the developer, for instance, to understand the data models and their context or to discover datasets with specific characteristics, such as temperature data from particular regions or datasets enriched with geospatial information. These capabilities enable the app to retrieve only the most relevant and contextually appropriate data for its use.

Monitoring UHIs requires fine-granular data to identify localized temperature variations effectively. In this paper, the app uses the city of Wuppertal's neighborhood categorization, a fine-grained subdivision of the city into smaller neighborhoods, based on the publicly available datasets provided by Wuppertal's open data platform. This categorization allows the system to associate temperature readings with specific districts, enabling a more precise analysis of urban heat patterns. By integrating data at the district level, the app ensures that insights are actionable and tailored to localized urban planning and public health interventions.

To enable secure and efficient data retrieval, the app employs Solid authentication via the `@inrupt/solid-client` library, ensuring authorized access to decentralized data storage. Upon authentication, datasets stored in Solid Pods are retrieved using the `getFile()` function, supporting CSV and JSON formats. The application processes these files by converting them into structured formats using a `parsing` library, enabling integration of multiple datasets. Data normalization ensures consistency, allowing dynamic visualization of temperature readings at the neighborhood level. By leveraging metadata descriptions and semantic models, the developer is able to find relevant datasets and understand the data structure to integrate them into one meaningful representation. The methodology was validated through systematic integration testing, focusing on authentication, file retrieval, and the processing of heterogeneous data sources. The tests ensured that data from diverse Solid Pods could be accurately retrieved, correctly parsed, semantically aligned, and reliably integrated into the Heat Monitoring App.

3.4. Evaluation Setting

To evaluate our Solid-based dataspace and our Heat Monitoring App, we constructed a first minimal experimental setting that replicates a real-world scenario involving decentralized data management and heterogeneous datasets. The Solid-based dataspace includes three server instances, each contributing and accessing data assets stored in diverse formats. The goal of the evaluation is to demonstrate the effective discovery, integration and utilization of statistical urban heat data using semantic technologies.

The evaluation setting comprises three data sources from a city, each representing temperature readings from district-level sensors with geolocation, but provided in different formats and using varying headers. Each of these data sources is thereby located on a different Solid server. The first dataset, in CSV format, includes attributes such as **city**, **QUARTIER**, **lat**, **lng**, **temp**, and **activated**. The second dataset, in JSON format, represents similar data with attributes such as **location**, **district**, **latitude**, **longitude**, **t**, and **activated**. The third dataset, also in CSV format, uses **c**, **q**, **lat**, **lng**, **t**, and **activation**. This heterogeneity demonstrates the challenges of integrating decentralized and inconsistent datasets, highlighting the necessity of semantic technologies when working with these datasets. Table 1 provides an overview of the different data sources used in the evaluation and highlights their structural differences.

Additionally, the dataspace incorporates three primary datasets while also including contributions from ten users, each maintaining a personal data Pod across three servers. While the presented datasets serve as the core data sources, the additional user-contributed data is not directly relevant for the

Table 1
Overview of Evaluation Datasets

Dataset Format	Attributes
CSV (dataset-1.csv)	`city`, `QUARTIER`, `lat`, `lng`, `temp`, `activated`
JSON (dataset-2.json)	`location`, `district`, `latitude`, `longitude`, `t`, `activated`
CSV (dataset-3.csv)	`c`, `q`, `lat`, `lng`, `t`, `activation`

application, demonstrating the nature of dataspaces, which are designed to accommodate a wide range of heterogeneous data.

4. Test Setting & Results

A simulated scenario was chosen as there is an inherent causality dilemma in the introduction of such systems. A fully operational dataspace must first exist to attract real-world data and users, but its adoption hinges on proving its functionality and value. Simulating the environment allows us to validate the core principles and interactions between components before real-world deployment.

The Heat Monitoring App is developed as part of the Gesundes Tal research project, which focuses on leveraging digital technologies to improve health and environmental resilience in urban areas. Based on apps like the urban heat monitoring, the project aims to provide actionable insights for policymakers, urban planners, and citizens, contributing to sustainable and data-driven decision-making.

Currently, the sensors used in the research project are being set up, and we are actively collaborating with the necessary departments to ensure that our simulated data accurately reflects real-world conditions. This close alignment between simulated and actual sensor data helps validate the feasibility of our approach while ensuring that the system remains adaptable to future real-world deployments.

4.1. Dataspace Structure

As highlighted in Section 3, the constructed Solid-based dataspace comprises three servers, each hosting multiple individual Solid Pods. Each server operates as an independent node within the dataspace, where users or institutions register their Pods to contribute and access data. While the data remains in its original format, such as JSON or CSV, semantic technologies are leveraged by underpinning it with RDF-based semantic models. This approach allows the system to retain the advantages of semantic interoperability and contextual understanding without requiring data conversion to RDF.

A CRUD-based mechanism facilitates the seamless addition of new Solid servers and Pods. New servers are registered within the data catalog, ensuring compliance with interoperability standards. Users can create, manage, and delete the information about their Pods through the interface of the data catalog.

4.2. Data Catalog

Solid servers and their associated Pods were registered manually within the catalog to ensure their metadata is indexed and accessible. The catalog supports the discovery of data, leveraging the DCAT standard to describe metadata.

To enhance the organization and discoverability of data stored in Solid Pods, we developed a semantically enriched data catalog (cf. Section 3). Participants from the dataspace can register their datasets with their corresponding semantic model from their Solid Pods using the interface of the catalog. We are capturing metadata for format, provenance, and access rights while linking them to RDF-based descriptions. This enables users and developers to efficiently search for and retrieve data through standardized interfaces.

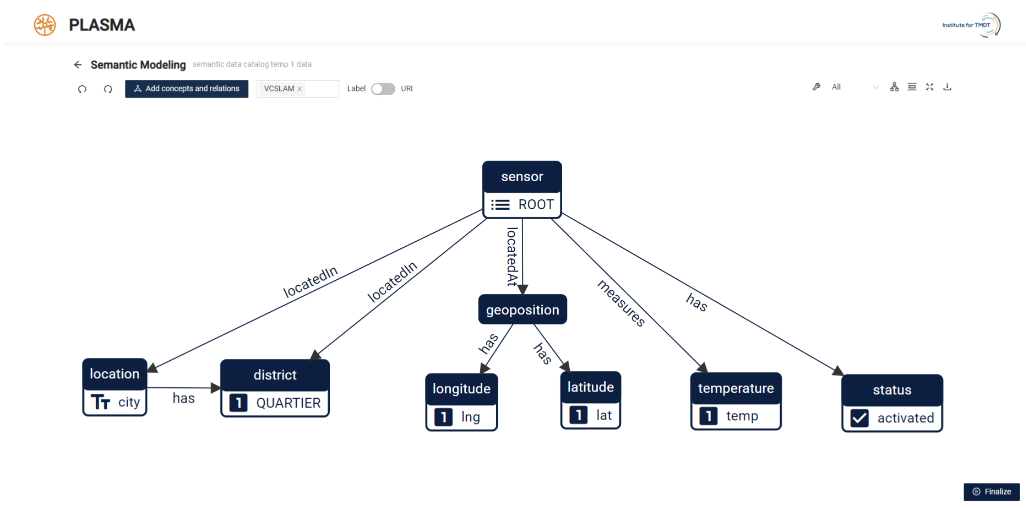

Figure 3: Semantic Model for the first temperature dataset.

4.3. Semantic Models

The semantic models for our test setting were created using PLASMA, a platform designed for the creation of semantic models, which is illustrated in Figure 3. Within PLASMA, we used the VC-Slam ontology [26] to semantically represent temperature data and its associated geospatial attributes, ensuring consistent and machine-readable descriptions.

4.4. Heat Monitoring App

The Heat Monitoring App was developed as a client application that integrates data sources from the Solid Dataspace. Its focus is on identifying and integrating temperature data from the city of Wuppertal. When the developer has found matching data with a certain criterion, he can retrieve the data and integrate it into the app's visualization. In terms of semantics, the use of DCAT ensures that the data format is clearly identified, while the semantic models enable the developer to determine that the data pertains to temperature observations from Wuppertal with geolocation.

The developed application successfully enables secure and efficient data retrieval from Solid Pods, leveraging Solid authentication and decentralized storage. Through systematic integration testing, authentication, file retrieval, and data processing workflows were validated, ensuring accurate and consistent integration of heterogeneous data sources. The Heat Monitoring App is currently capable of processing CSV and JSON files into a unified internal data model, normalizing heterogeneous inputs for seamless aggregation and visualization.

One of the key challenges is integrating heterogeneous datasets into the Heat Monitoring App. The semantic models provide information on how the individual files are structured. Based on these information, the developer can then align the different data schemas, particularly for temperature values, which were represented using varying attribute names and units across datasets. For instance, some datasets used `temp` to denote temperature readings, while others used `t` or `temperature_value`.

To address these inconsistencies, a first simple syntactic schema alignment mechanism was implemented within the Heat Monitoring App. This mechanism leverages semantic annotations in the semantic models to normalize attribute names. By mapping dataset attributes to a unified ontology — specifically, the VC-Slam ontology — the system ensures that all temperature values are interpreted consistently regardless of their original schema. The alignment process is illustrated in Figure 4, showcasing how disparate data sources can be harmonized before integration into the Heat Monitoring App.

This schema alignment approach plays a crucial role in enabling seamless data integration, ensuring that datasets with varying structures can be automatically processed without manual intervention. It

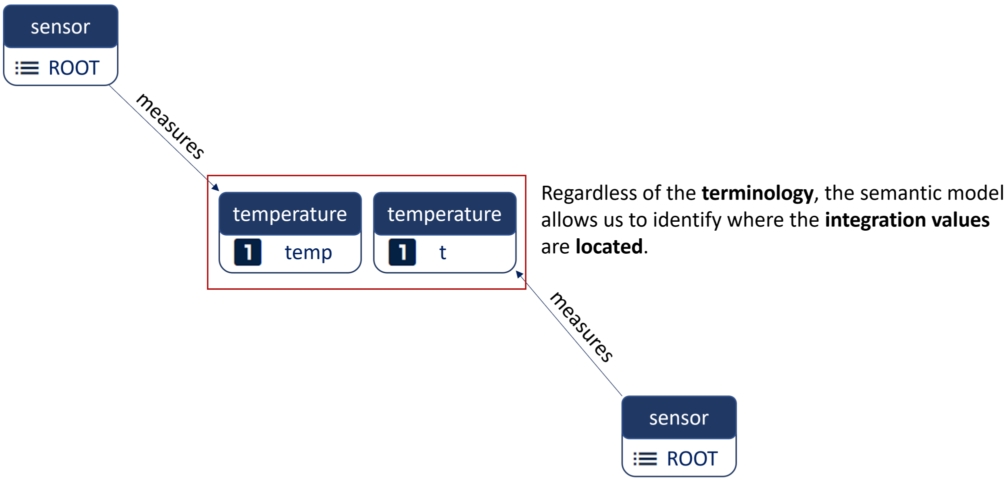

Figure 4: Schema alignment for integrating heterogeneous temperature datasets into the Heat Monitoring App.

enhances interoperability within the Solid-based dataspace, allowing diverse temperature datasets to be leveraged effectively for urban heat monitoring.

However, this approach is currently limited to simple schemas where mappings remain relatively straightforward. In cases involving, e.g., nested file structures, it falls short, as hierarchical relationships and complex dependencies require more sophisticated strategies. Currently, most alignment relies on manually identifying corresponding field names and implementing normalization mechanisms. However, our goal is to automate this process further in the future. By incorporating automated techniques for identifying relevant datasets and fully automated mechanisms for schema alignment shall automate seamless dataset integration in app development for tasks like these. Overcoming these challenges remains therefore a key direction for future research.

Although the integration of private datasets within the Solid-based dataspace is possible, our app can currently only interact with public data, which is a further limitation of this paper. While the app demonstrates data discovery and integration, it operates under the assumption that data is publicly available. In a fully functional dataspace, private data should also be discoverable and requestable. The correct sequence would involve the app identifying private data, sending a request to the data owner for access and, upon approval, integrating the data into its workflows. Data owners would retain the ability to revoke access rights, causing the data to disappear from the app. This dynamic interaction, while not yet implemented, underscores the potential of the proposed system to balance privacy and utility in real-world applications.

This proof of concept highlights contributions and added value, especially in the context of addressing UHIs. The integration of an initial data catalog for Solid enriched with semantic models complements existing approaches by enabling meaningful descriptions of data assets. The showcase of the data catalog is illustrated in Figure 5. By employing the VC-Slam ontology, the system adheres to widely recognized standards, promoting interoperability and reuse. The current focus on statistical batch data showcases its utility for retrospective analyses, although it is limited by its inability to handle streaming data effectively.

In this context, it is worth noting advancements in RDF stream processing, such as the concept of Stream Containers proposed by Schraudner and Harth [27], which offer a scalable and web-aligned approach to streaming data integration. These advancements could enhance the system's ability to handle real-time data streams, bridging the current gap between batch processing and dynamic, continuous data integration.

As a result, visualizations of temperature readings at the neighborhood level were achieved as illustrated in Figure 1. This demonstrates the practical applicability of the approach and enables the identification of UHIs through the developed application.

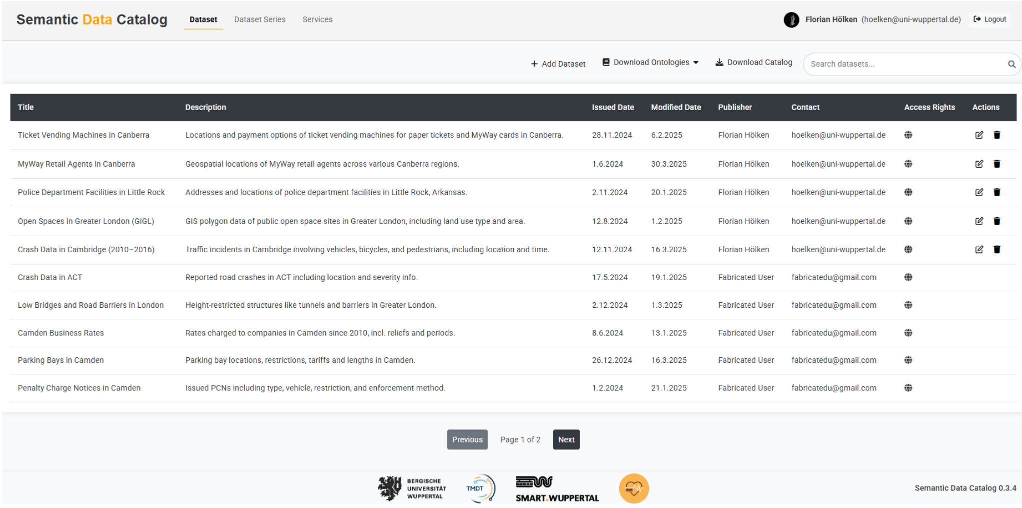

Figure 5: Showcase of the data catalog, illustrating DCAT metadata enriched datasets.

5. Conclusion

This paper presented a novel approach to urban heat island (UHI) monitoring by leveraging a decentralized Solid-based dataspace enriched with a data catalog and semantic modeling. Through the development of a Heat Monitoring App, we demonstrated the feasibility of interoperable, user-centric urban climate data management. Our approach addresses key challenges of urban data ecosystems, promoting data sovereignty, interoperability, and contextual understanding via semantic modeling and Linked Data principles.

However, several limitations remain. The system primarily relies on batch data, limiting its capability for real-time environmental monitoring. Although semantic enrichment facilitates integration, schema alignment still requires significant manual effort, hindering scalability. Furthermore, the approach assumes that datasets are publicly available or accessible via predefined permissions, lacking mechanisms for dynamic access negotiation. Ensuring data consistency and reliability across multiple distributed Pods remains challenging, and currently, only temperature data with geospatial information is considered, restricting broader environmental analysis. These areas highlight important directions for future improvements to enhance the robustness, scalability, and applicability of the proposed approach in real-world urban settings.

6. Future Work

Future research will address these limitations along several directions. First, we aim to integrate real-time data streams by leveraging RDF Stream Processing (RSP) techniques, allowing continuous ingestion of temperature and environmental data from IoT sensors and weather stations. Second, enhancing automation in schema alignment and semantic enrichment—potentially using Large Language Models (LLM) will reduce manual effort and increase scalability. Third, managing access to private datasets while preserving data sovereignty remains a key challenge. To address this, we will explore privacy preserving access control mechanisms within the Solid framework, enabling dynamic data request and negotiation functionalities. Finally, optimizing the scalability, system performance, and user experience through improved interfaces, intuitive visualization tools, and active stakeholder engagement will be essential for successful deployment in larger urban settings.

By addressing these challenges, future research aims to bridge the gap between theoretical advancements in dataspaces and their practical application in urban climate resilience, ultimately fostering a more decentralized, participatory, and data-driven approach to mitigating urban heat islands.

Acknowledgements

This work has been supported as part of the research project *Gesundes Tal* in collaboration with the city of Wuppertal, funded by the Federal Ministry of Housing, Urban Development and Building (BMWSB) and the Reconstruction Loan Corporation (KfW) through the funding program “Modellprojekte Smart Cities: Stadtentwicklung und Digitalisierung” (grant number 19454890).

Declaration on Generative AI

During the writing of this paper, the author(s) used DeepL and GPT-4o in order to: Grammar, translation and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers, Big data: The next frontier for innovation, competition, and productivity, 2011.

[2] M. Hupperz, A. Gieß, The interplay of data-driven organizations and data spaces: Unlocking capabilities for transforming organizations in the era of data spaces, 2024. doi:10.24251/HICSS.2024.541.

[3] S. Meckler, R. Dorsch, D. Henselmann, A. Harth, The web and linked data as a solid foundation for dataspaces, in: Companion Proceedings of the ACM Web Conference 2023, WWW ’23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 1440–1446. URL: https://doi.org/10.1145/3543873.3587616. doi:10.1145/3543873.3587616.

[4] T. Kastner, C. H.-J. Braun, A. Both, D. Yeboah, S. J. Schmid, D. Schraudner, T. Käfer, A. Harth, Data-sovereign enterprise collaboration using the solid protocol, in: 20th International Conference on Semantic Systems (SEMANTICS’24), Amsterdam, 17th-19th September 2024, volume 3759 of *CEUR workshop proceedings*, CEUR-WS, 2024, pp. Art.–Nr.: 10.

[5] T. R. Oke, The energetic basis of the urban heat island, Quarterly Journal of the Royal Meteorological Society 108 (1982) 1–24. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49710845502. doi:https://doi.org/10.1002/qj.49710845502. arXiv:https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49710845502.

[6] M. Santamouris, Analyzing the heat island magnitude and characteristics in one hundred asian and australian cities and regions, Science of The Total Environment 512-513 (2015) 582–598. URL: https://www.sciencedirect.com/science/article/pii/S0048969715000753. doi:https://doi.org/10.1016/j.scitotenv.2015.01.060.

[7] Intergovernmental Panel on Climate Change (IPCC), Climate Change 2021: The Physical Science Basis, Cambridge University Press, 2021. URL: https://www.ipcc.ch/report/ar6/wg1/.

[8] I. Stewart, T. Oke, Local climate zones for urban temperature studies, Bulletin of the American Meteorological Society 93 (2012) 1879–1900. doi:10.1175/BAMS-D-11-00019.1.

[9] M. Franklin, A. Halevy, D. Maier, From databases to dataspaces: a new abstraction for information management, SIGMOD Rec. 34 (2005) 27–33. URL: https://doi.org/10.1145/1107499.1107502. doi:10.1145/1107499.1107502.

[10] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to european digital sovereignty with gaia-x and idsa, IEEE Network 35 (2021) 4–5. doi:10.1109/MNET.2021.9387709.

[11] I. D. S. Association, IDSA Position Paper: GAIA-X and IDS, Technical Report, International Data Spaces Association, 2020. URL: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf, accessed: 2024-12-19.

[12] T. Berners-Lee, Linked data - design issues, https://www.w3.org/DesignIssues/LinkedData, 2006. Accessed: 2025-04-28.

[13] E. Curry, Dataspaces: Fundamentals, Principles, and Techniques, Springer International Publishing, Cham, 2020, pp. 45–62. URL: https://doi.org/10.1007/978-3-030-29665-0_3. doi:10.1007/978-3-030-29665-0_3.
[14] B. Vu, C. Knoblock, J. Pujara, Learning semantic models of data sources using probabilistic graphical models, 2019, pp. 1944–1953. doi:10.1145/3308558.3313711.
[15] M. Taheriyan, C. A. Knoblock, P. Szekely, J. L. Ambite, A scalable approach to learn semantic models of structured sources, in: 2014 IEEE International Conference on Semantic Computing, 2014, pp. 183–190. doi:10.1109/ICSC.2014.13.
[16] A. Paulus, A. Burgdorf, A. Pomp, T. Meisen, Recent advances and future challenges of semantic modeling, in: 2021 IEEE 15th International Conference on Semantic Computing (ICSC), 2021, pp. 70–75. doi:10.1109/ICSC50631.2021.00016.
[17] A. Pomp, A. Paulus, D. Klischies, C. Schwier, T. Meisen, A web-based ui to enable semantic modeling for everyone, Procedia Computer Science 137 (2018) 249–254. URL: https://www.sciencedirect.com/science/article/pii/S1877050918316296. doi:https://doi.org/10.1016/j.procs.2018.09.024, proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria.
[18] C. A. Knoblock, P. Szekely, J. L. Ambite, A. Goel, S. Gupta, K. Lerman, M. Muslea, M. Taheriyan, P. Mallick, Semi-automatically mapping structured sources into the semantic web, in: E. Simperl, P. Cimiano, A. Polleres, O. Corcho, V. Presutti (Eds.), The Semantic Web: Research and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 375–390.
[19] A. Paulus, A. Burgdorf, T. Langer, A. Pomp, T. Meisen, S. Pol, Plasma: A semantic modeling tool for domain experts, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 4946–4950. URL: https://doi.org/10.1145/3511808.3557184. doi:10.1145/3511808.3557184.
[20] M. Taheriyan, C. Knoblock, P. Szekely, J. L. Ambite, Leveraging linked data to discover semantic relations within data sources, volume 9981, 2016, pp. 549–565. doi:10.1007/978-3-319-46523-4_33.
[21] S. Vande Walle, Data spaces, data ecosystems and competition law: A spotlight on the eu's gaia-x, catena-x and cofinity-x projects, 2024.
[22] I. D. S. Association, IDS Reference Architecture Model 3.0, International Data Spaces Association, 2019. URL: https://internationaldataspaces.org/wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf, accessed: 2023-12-11.
[23] G. Schimak, L. Petronzio, T. Pariente Lobo, Approaching cross-domain search in environmental applications - towards linked data, volume 413, 2013. doi:10.1007/978-3-642-41151-9_8.
[24] S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, Big data in smart farming – a review, Agricultural Systems 153 (2017) 69–80. URL: https://www.sciencedirect.com/science/article/pii/S0308521X16303754. doi:https://doi.org/10.1016/j.agsy.2017.01.023.
[25] M. Taheriyan, C. A. Knoblock, P. Szekely, J. L. Ambite, Learning the semantics of structured data sources, Journal of Web Semantics 37-38 (2016) 152–169. URL: https://www.sciencedirect.com/science/article/pii/S1570826815001444. doi:https://doi.org/10.1016/j.websem.2015.12.003.
[26] A. Burgdorf, A. Paulus, A. Pomp, T. Meisen, VC-SLAM—A Handcrafted Data Corpus for the Construction of Semantic Models, Data 7 (2022) 17. URL: https://doi.org/10.3390/data7020017. doi:10.3390/data7020017.
[27] D. Schraudner, A. Harth, Stream containers for resource-oriented rdf stream processing, 2022. doi:10.48550/arXiv.2202.13630.