

The Role of RDF Stream Processing in an Smart City ICT Infrastructure - The Aspern Smart City Use Case

Josiane Xavier Parreira^(✉), Deepak Dhungana,
and Gerhard Engelbrecht

Siemens AG Österreich, Vienna, Austria

{josiane.parreira,deepak.dhungana,gerhard.engelbrecht}@siemens.com

Abstract. In this paper we discuss the opportunities of adopting RDF stream processing in the context of smart cities. As a concrete example we take the Aspern Smart City Research project – one of the largest smart city projects in Europe – which aims at overcoming silos in smart grid and smart building domains. We present the envisioned smart ICT infrastructure and identify how RDF Stream processing can be explored in the different interactions among data sources, storage centers and applications/services.

Keywords: RDF stream processing · Semantic web · Smart cities · Aspern urban lakeside district

1 Introduction

The Aspern Smart City Research (ASCR)¹, is a joint initiative between Siemens AG Österreich, Wiener Netze (a Power Grid operator), Wien Energie (an energy supplier), Wirtschaftsagentur Wien, and Wien 3420 Aspern Development AG. Started in October 2013, with an expected duration of 5 years and a budget of almost 40 million Euros, the project will have as testbed a “living laboratory” which is being created in the urban-lake-side district of Aspern, one of the largest urban development projects in Europe². This area will include apartments, offices, and a business, science, research, and education quarter. Altogether, it will cover around 240 hectares. Fifty percent of the space is reserved for public areas – plazas, parks, and recreation areas. Step by step, between now and 2030, the district will evolve into an intelligent city of the future, with 20,000 residents and 20,000 additional jobs.

The Aspern project represents an opportunity to develop a long-term integrated concept for an energy-optimized city district, using appropriate technologies, products, and solutions in a real-world infrastructure. The overall goal is to make the whole system “smarter”, by having power supplies, building systems,

¹ <http://www.ascr.at/>.

² <http://www.aspern-seestadt.at/en>.

intelligent power grids, and information and communication technologies (ICT) interacting in an optimal manner. For example, part of the project involves connecting buildings that have different functions, i.e. offices and apartments, to the low-voltage distribution network. This will allow efficient management of the energy exchange between buildings and optimization of the local energy consumption. This offers building operators the possibility to participate actively on the energy markets. To this end, data from the different domains will be combined and used by different applications and services.

The project envisions a Smart ICT platform at the core of the interaction between the different players in a smart city. The Smart ICT platform is responsible for, among other things, mediating the interaction between data owners/publishers and applications/services. Data owners can make their data available via the ICT platform – either directly or via a data publisher – which are then stored locally. Application and services developers can access the stored data via the platform to create new applications, which in turn can be made available to end users also via the platform. The ICT platform supports different other functionality such as access control, policy enforcement, billing, monitoring and discovery.

We have identified different opportunities in the Smart ICT platform in which RDF Stream processing can aid delivering efficient and scalable data integration and analytics solutions for heterogeneous domains. For example, RDF Stream processing can be integrated in existing ETL tools to deliver semantically annotated and aggregated information from the raw data sources, which is then stored and combined with existing knowledge bases (e.g. buildings' layouts), before being delivered to third party applications and services. Real time log data from applications/services can, in turn, be analyzed to optimize the data acquisition process.

Next we describe the ASCR's infrastructure and its main components, in particular the Smart ICT platform. We then identify where RDF Stream processing can be explored by the Smart ICT platform in the different interactions among data sources, storage and applications/services, followed by a deeper discussion about its benefits.

2 The ASCR Infrastructure

The overall goal of Aspern is to deliver “smarter” solutions for energy utilization, by having power supplies, building systems, intelligent power grids, and information and communication technologies (ICT) interacting in an optimal manner. To this end, the ASCR infrastructure is responsible for the integration, interaction, analysis and provisioning of data coming from smart grids and smart buildings (e.g. temperature, energy consumption, water consumption, power demand), as well as external data sources (e.g. weather, city events, energy market, traffic reports). It is clear that the success of the infrastructure will depend in efficient solutions for handling heterogeneous data. The combination of data from different domains will not only lead to better forecast models, but it will also enable

exploratory analysis to discover new correlations among the data, thus improving even further the optimization measures. In addition, the infrastructure also serves as a tool to aggregate and compare data at different levels, e.g., whole city, districts and building complexes.

Access to the data and analytical tools is limited to grids and building management systems. Smart citizens should be able to access data relevant to them and to also contribute by providing services to other users. For that, they will take the role of application developers, and applications can be made available via the ICT infrastructure. Figure 1 shows the ASCR Infrastructure and its main components.

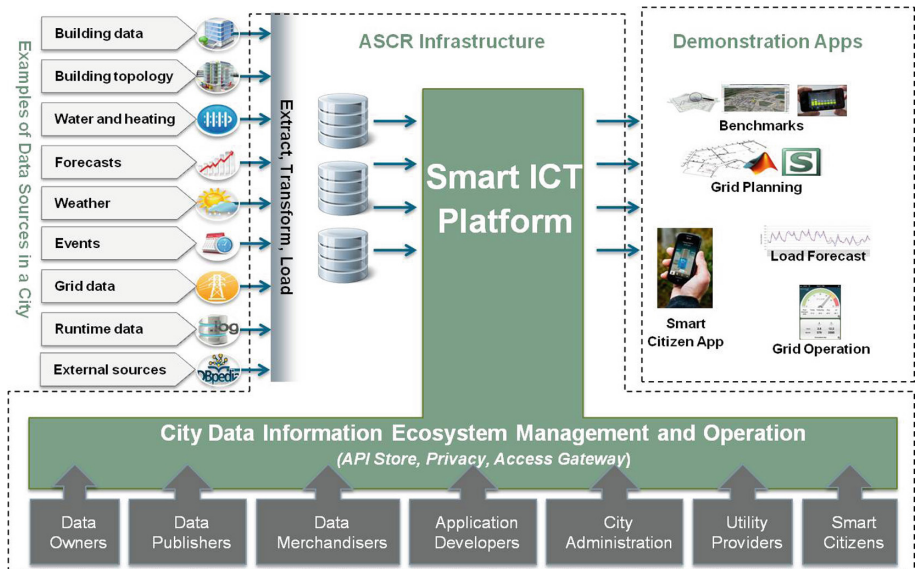


Fig. 1. The ICT infrastructure as a backbone for applications in Aspern [9]

Data from different sources in the city go over Extract, Transformation, Load (ETL) processes before being stored locally. The Smart ICT platform then performs different data analyses in the stored data, for instance [9]:

- **Benchmarking** of the different aspects of the ecosystem, in order to assess the performance of the optimization measures in the buildings and grid operation.
- **Load forecast**, by improving existing models with additional data from external sources.
- **Grid planning**, by early detection of anomalies or threshold violations in the low voltage network and identification of possible causes.

The platform delivers both the data and the data analysis to different applications, in order to provide the smart citizens with useful services. For example, mobile applications can be developed to give energy saving advices to citizens, based on their energy consumption, preferences, schedule and other external factors (e.g. weather). Citizens are encouraged to actively contribute to the platform, by developing applications on top of the available data and services. Allowing data access to third party users fosters a rapid increase of the number of functionalities offered by the platform, thus increasing its uptake they the community.

Different stakeholders are part of this complex ecosystem – data owners, data publishers, smart citizens, city administrators, etc. – and they all have different profiles and needs. Therefore, mechanisms for identity management, authentication, access control and policy enforcement are in place to ensure that users have the right credentials and the right subscription plan, according to the data source(s). In addition, an API store is in place to enable users to browse and discover the available applications and services.

3 RDF Stream Processing for Smart ICT Interactions

Extensive work in the past decade has shown how Semantic Web and Linked Data technologies can aid overcoming data silos [2,6,11,23]. Recently, the Semantic Web concepts have been extended to streaming information [7,15,18,19,22]. As RDF (Resource Description Framework) is the de facto standard for semantic data representation, it was expected that semantic streams would follow the same pattern, thus leading to many efforts towards RDF Stream processing (RSP) [1,5,8,13]. RDF Stream processing enables data integration – not only among heterogeneous stream data sources, but also with other, possibly heterogeneous, existing sources. The advent of concepts such as of Internet of Things [24], Web of Things [20], and Industry 4.0 [16] – where integration of sensory information is at the core – also make RSP a timely topic with increasing attention by researchers and developers.

The RSP paradigm is well aligned with the requirements of the ASCR Smart ICT platform. A lot of the data that goes into the platform comes from sensor sources, for example, energy meters, room occupancy, temperature, and they need to be integrated with other knowledge bases, like building layout and grid topology. Moreover, data stored in the platform needs to be deliver to the applications on demand, and its often the case that they are streamed into the different applications. At last, application log streams are fed into the system for monitoring and performance improvement.

Based on this analysis, we have identified three different stages of the Smart ICT data pipeline, as shown in Fig. 2 – from data sources to local storage, from local storage to applications, and from applications back to data sources. RDF Stream processing can be explored in all these stages, in different ways. These are discussed in detail in the following sections. For some cases, existing work in RDF Stream processing tools and techniques already fulfills the requirements, while some are still open challenges for the RSP community.

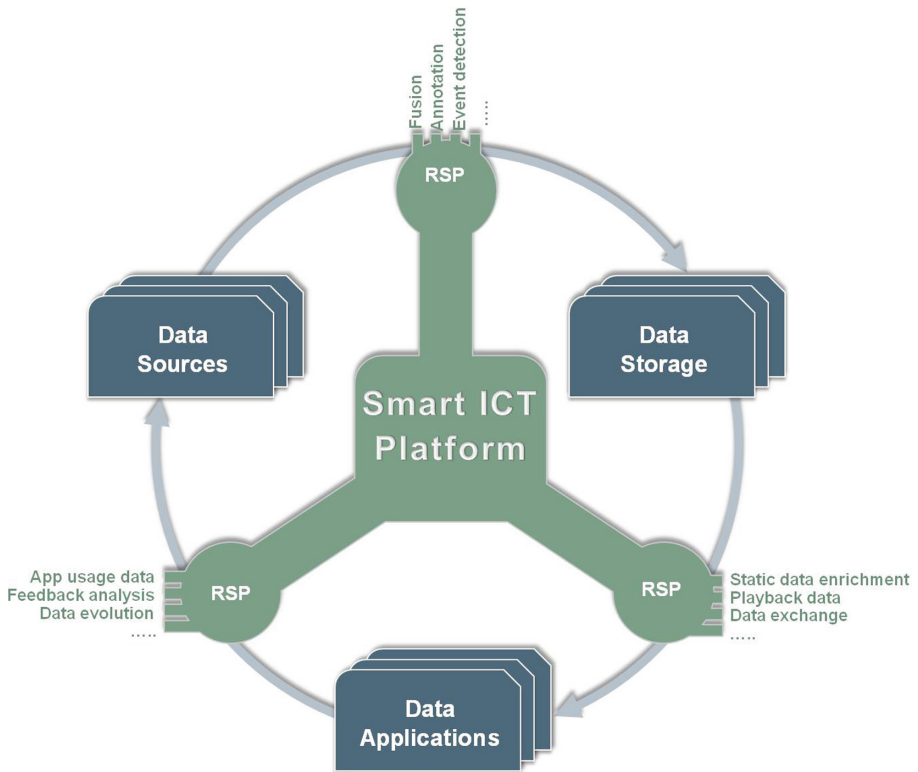


Fig. 2. RDF stream processing in the smart ICT platform.

3.1 From Data Sources to Data Storage

Heterogeneous data streams from diverse application domains (e.g., energy, traffic, event calendars, and environmental sensors for pollution or weather warnings, GIS databases) are at the heart of any smart city – as the data from these sources and their timely analysis can highly impact the smartness of the city. While data integration is paramount, not all data sources and not all data from these sources are relevant for decision makers, or citizens of the city. The dilemma for data-infrastructure engineers and data scientists is clear: what part of the data should be stored, how it should be stored and for how long? Here, RDF stream processing tools and techniques can be used to continuously monitor the data sources and perform pre-processing, filtering and integration, in order to store only the relevant parts of the data streams. For example, in the case of the Aspern project, RSP tools and techniques could be used for:

- **Data annotation:** Different RSP platforms, such as the Linked Stream Middleware [15], Graph of Things³ and BOTTARI [4], provide the so-called wrap-

³ <http://graphofthings.org/>.

pers, which can take different input formats and produce semantically annotated content. This is the first step to enable semantic integration and it can be tailored to capture different relevant information that goes beyond the measures themselves, such as provenance and accuracy. In Aspern, information about ownership is an example of an important data feature that should be kept through all the data pipeline, and RDF Streams can easily provide modeling abstractions that fits the different requirements. Existing wrappers still require initial manual configuration so that the input is correctly mapped to the semantic description. Automatic semantic annotation of stream data is still an open challenge.

- **Stream storage:** The platforms mentioned in the item above also support storing and querying historical stream data. Efficient encoding solutions for RDF already exist [10,17] to reduce the storage size, while still allowing a limited set of query operations. These could be extended to deal with the temporal aspect of RDF Streams. In addition, in cases where there is no need to store the original input data, RSP engines can provide data filtering and aggregation which can potentially lead to storage savings.
- **Data fusion:** RSP engines offer operators and functions similar to those found in relational database systems (RDBS), such as joins, unions, and aggregation, therefore supporting data fusion. Solutions such as C-SPARQL [5], SPARQL_{Stream} [8], CQELS [13] and EP-SPARQL [1] have been extensively benchmark and results can be found at [14,25]. The benchmark frameworks can aid the choice of the most adequate RSP engine, given a specific set of requirements.
- **Event detection:** A large number services within Aspern relies on event detection, for example, grid planning. While most of the RSP engines are based on the Stream Data Management Systems (DSMS) paradigm, some of the existing work, such as EP-SPARQL, takes an Complex Event Processing (CEP) approach. EP-SPARQL's support of Allens temporal relationships enables event detection on RDF streams. RDF Stream Reasoning is also gained popularity [21]. While a few works already support it, further research is still needed to address scalability issues due to the complexity of reasoning tasks.

3.2 From Data Storage to Data Applications

One of the key issues with streaming data sources is that the data may not be available when it is needed, as sensors might be faulty. Many applications relying on streaming data sources however assume that the data is always accessible and available. In order to deal with this issue, the Aspern project stores the data locally in a big data processing infrastructure and passes these to the applications as required in a controlled manner. Nevertheless, many applications still expect streaming data, simply due to the very nature of some data points in the city. In order to deal with this, RSP tools and techniques could be used. In the Aspern project, this use case has many facets, which gives several opportunities to apply RSP.

- **Playback data to applications:** Most of RSP solutions follow the query semantics defined by the Continuous Query Language (CQL) [3]. One class of operators is the “relation-to-stream” operators that produce a stream from a relation. That enables stored data to be delivered to application as streams.
- **Integration of static information:** Similar to the data fusion task described in Sect. 3.1, data integration is one of the main reasons for adopting RSP in the Aspern project. As mentioned earlier, Semantic Web can help overcoming silos, by promising seamless way of integration data. The benefits of using semantic technologies go beyond the ASCR project and can further facilitate the integration of other city data providers. Furthermore, it can easily be extended to other applications where heterogeneous devices must cooperate, for example, in the Industry 4.0 vision. All the solutions mentioned in Sect. 3.1 can be explored for data integration tasks.
- **Semantic streams as a data exchange format:** Different serializations formats available for RDF are also supported by RSP engines, such as Turtle⁴, N-triples⁵, JSON-LD⁶. However, they might not be suitable for cases involving constrained devices or when the communication is over channels with low transfer rates, for example, mobile applications communicating via 3G networks. Binary representations, similar to the ones used for compressing, and compact representations (e.g. EXI [12]) are currently being considered by ongoing research.

3.3 From Data Applications to Data Sources

Different applications used by the citizens and decision makers in the context of a smart city are themselves generators of data. They represent a very valuable data sources, as they can give a lot of insights into how the applications are being used, whether the applications themselves are of any value to the users and how these can be improved. The events generated by the applications, the data browsing “behavior patterns” of the users, and the direct feedback that comes from the users, annotated with the contextual information – all of these data points come as streaming sources– which can be treated as yet another data sources in the smart city context. In the context of the Aspern project, such data could be collected, analyzed and stored using RDF stream processing tools and techniques.

- **Feedback analysis:** Data analysis in general is crucial in many applications, for example benchmarking and load forecast, and as such it is not limited to feedback data logs. Nevertheless, feedback analysis concerns the overall performance of the platform, and it therefore deserves special attention. How the performance is perceived is highly dependent of the user needs and usage behavior. Therefore, we expect different user groups to behave differently (e.g.

⁴ <http://www.w3.org/TR/turtle/>.

⁵ <http://www.w3.org/TR/n-triples/>.

⁶ <http://www.w3.org/TR/json-ld/>.

data publishers vs. app developers). By identifying these patterns we would be able to identify groups of users and automatically adapt the platform to their needs. Feedback analysis is also important to monitor and predict the performance of the overall infrastructure in order to maintain high performance and reduce down times.

Feedback analysis can be quite complex and might involved a number of steps. Complex event processing and reasoning over RDF streams can provide the first step for identifying meaningful patterns. An interesting, still unsolved challenge, would be how to combined RSP methods with traditional machine learning approaches to enable dynamic and complex data analysis.

4 Conclusion

Semantic Web technologies provide a schema-free data abstraction that fosters data reuse and integration, across different domains. RDF Stream Processing brings the Semantic Web paradigm to streaming information, thus bridging the gap between stream resources and knowledge bases. The Aspern project represents an opportunity to apply RSP solutions in a real-world infrastructure with the goal of developing a long-term integrated concept for an energy-optimized city district. In this paper we have consider the Smart ICT infrastructure from Aspern to demonstrate the benefits of RSP. We have looked at concrete functionalities in the platform and discussed how RSP can be applied.

Despite being a recent topic, research on RDF Stream Processing has already made considerable progress towards supporting tools and methods. Nevertheless there are still a large number of open challenges.

One very important step to foster the development and uptake of RSP technologies was the creation of the W3C RDF Stream Processing community group⁷. In this group, which was created in 2013, researchers in the area of RSP have joint forces to define common models for producing, transmitting and continuously querying RDF Streams. The expected outcome is a set of specifications for RDF extensions for streaming data – including modeling, querying, syntax, semantics, and service interfaces – to be adopted by future solutions.

References

1. Anicic, D., Fodor, P., Rudolph, S., Stojanovic, N.: EP-SPARQL: a unified language for event processing and stream reasoning. In: WWW, pp. 635–644. ACM (2011)
2. Antoniou, G., van Harmelen, F.: A Semantic Web Primer, 2nd edn. The MIT Press, Cambridge (2008)
3. Arasu, A., Babu, S., Widom, J.: The CQL continuous query language: semantic foundations and query execution. *VLDB J.* **15**(2), 121–142 (2006)
4. Balduini, M., Celino, I., DellAglio, D., Valle, E.D., Huang, Y., Lee, T., Kim, S.-H., Tresp, V.: BOTTARI: an augmented reality mobile application to deliver personalized and location-based recommendations by continuous analysis of social media streams. *Web Semant. Sci. Serv. Agents World Wide Web* **16**(5), 33–41 (2012)

⁷ <https://www.w3.org/community/rsp/>.

5. Barbieri, D.F., Braga, D., Ceri, S., Valle, E.D., Grossniklaus, M.: C-SPARQL: a continuous query language for RDF data streams. *Int. J. Semant. Comput.* **4**(1), 3–25 (2010)
6. Bizer, C., Heath, T., Berners-lee, T.: Linked data - the story so far. *Int. J. Semant. Web Inf. Syst.* **5**, 1–22 (2009)
7. Bouillet, E., Feblowitz, M., Liu, Z., Ranganathan, A., Riabov, A.V., Ye, F.: A semantics-based middleware for utilizing heterogeneous sensor networks. In: Aspnes, J., Scheideler, C., Arora, A., Madden, S. (eds.) *DCOSS 2007*. LNCS, vol. 4549, pp. 174–188. Springer, Heidelberg (2007)
8. Calbimonte, J.-P., Corcho, O., Gray, A.J.G.: Enabling ontology-based access to streaming data sources. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) *ISWC 2010, Part I*. LNCS, vol. 6496, pp. 96–111. Springer, Heidelberg (2010)
9. Engelbrecht, G., Dhungana, D., Parreira, J.X.: Smart ICT in der seestadt aspern (poster). In: *Smart Grids Week* (2015)
10. Arias Gallego, M., Corcho, O., Fernández, J.D., Martínez-Prieto, M.A., Suárez-Figueroa, M.C.: Compressing semantic metadata for efficient multimedia retrieval. In: Bielza, C., Salmerón, A., Alonso-Betanzos, A., Hidalgo, J.I., Martínez, L., Troncoso, A., Corchado, E., Corchado, J.M. (eds.) *CAEPIA 2013*. LNCS, vol. 8109, pp. 12–21. Springer, Heidelberg (2013)
11. Hitzler, P., Krötzsch, M., Rudolph, S.: *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC, London (2009)
12. Käbisch, S., Peintner, D., Anicic, D.: Standardized and efficient RDF encoding for constrained embedded networks. In: Gandon, F., Sabou, M., Sack, H., d’Amato, C., Cudré-Mauroux, P., Zimmermann, A. (eds.) *ESWC 2015*. LNCS, vol. 9088, pp. 437–452. Springer, Heidelberg (2015)
13. Le-Phuoc, D., Dao-Tran, M., Xavier Parreira, J., Hauswirth, M.: A native and adaptive approach for unified processing of linked streams and linked data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) *ISWC 2011, Part I*. LNCS, vol. 7031, pp. 370–388. Springer, Heidelberg (2011)
14. Le-Phuoc, D., Dao-Tran, M., Pham, M.-D., Boncz, P., Eiter, T., Fink, M.: Linked stream data processing engines: facts and figures. In: Cudré-Mauroux, P., et al. (eds.) *ISWC 2012, Part II*. LNCS, vol. 7650, pp. 300–312. Springer, Heidelberg (2012)
15. Le-Phuoc, D., Nguyen-Mau, H.-Q., Parreira, J.-X., Hauswirth, M.: A middleware framework for scalable management of linked streams. *Web Semant. Sci. Serv. Agents World Wide Web* **16**(5), 42–51 (2012)
16. Lee, J., Lapira, E., Bagheri, B., Kao, H.: Recent advances and trends in predictive manufacturing systems in big data environment. *Manuf. Lett.* **1**(1), 38–41 (2013)
17. Owens, A., Seaborne, A., Gibbins, N., Schraefel, m.c.: A clustered triple store for jena, *Clustered TDB* (2008)
18. Sequeda, J.F., Corcho, O.: Linked stream data: a position paper. In: *SSN* (2009)
19. Sheth, A., Henson, C., Sahoo, S.S.: Semantic sensor web. *IEEE Internet Comput.* **12**, 78–83 (2008)
20. Uckelmann, D., Harrison, M., Michahelles, F.: *Architecting the Internet of Things*, 1st edn. Springer, Heidelberg (2011)
21. Valle, E.-D., Ceri, S., van Harmelen, F., Fensel, D.: It’s a streaming world! reasoning upon rapidly changing information. *IEEE Intell. Syst.* **24**, 83–89 (2009)

22. Whitehouse, K., Zhao, F., Liu, J.: Semantic streams: a framework for composable semantic interpretation of sensor data. In: Römer, K., Karl, H., Mattern, F. (eds.) EWSN 2006. LNCS, vol. 3868, pp. 5–20. Springer, Heidelberg (2006)
23. Wood, D.: Linking Enterprise Data, 1st edn. Springer, New York (2010)
24. Zanella, A., Bui, N., Castellani, A.-P., Vangelista, L., Zorzi, M.: Internet of things for smart cities. *IEEE Internet Things J.* **1**(1), 22–32 (2014)
25. Zhang, Y., Duc, P.M., Corcho, O., Calbimonte, J.-P.: SRBench: a streaming RDF/SPARQL benchmark. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012, Part I. LNCS, vol. 7649, pp. 641–657. Springer, Heidelberg (2012)