

# System of Systems Information Interoperability using a Linked Dataspace

Edward Curry

Digital Enterprise Research Institute,  
National University of Ireland, Galway.  
ed.curry@deri.org

**Abstract** - *System of Systems pose significant technical challenges in terms of information interoperability that require overcoming conceptual barriers (both syntax and semantic) and technological barriers. This paper presents an approach to System of Systems information interoperability based on the Dataspace data management abstraction and the Linked Data approach to sharing information on the web. The paper describes the fundamentals of the approach and demonstrates the concept with a System of Systems for enterprise energy management.*

**Keywords:** System of systems, interoperability, linked data, energy management.

## 1 Introduction

Emerging next generation smart environments such as Smart Grids, Smart Cities, and Smart Enterprises are complex systems that require a complete and holistic knowledge of their operations for effective decision-making. Multiple information systems currently operate within these environments and real-time decision support will require a System of Systems (SoS) approach to provide a functional view of the entire environment to understand, optimize, and reinvent processes. The required system of systems will need to connect systems that cross-organizational boundaries, come from multiple domains, (i.e. finance, manufacturing, facilities, IT, water, traffic, waste, etc.) and operate at different levels (i.e. region, district, neighborhood, building, business function, individual). These SoS pose many significant challenges, including the need for flexible mechanisms for information interoperability.

This paper presents an approach to SoS information interoperability based on two emerging trends, Dataspace and Linked Data. A Dataspace is a conceptual approach to information management that supports the co-existence of heterogeneous data with an incremental approach to interoperability. Linked data is a technology that leverages the web architecture to share data in a flexible and incremental manner to reduce technological barriers. When used together the resulting Linked Dataspace can provide a viable approach to SoS information interoperability. This paper discusses the approach with section 2 examining SoS interoperability requirements; section 3 discussing dataspace and linked data. Section 4 details the role of a

linked dataspace for SoS interoperability, with section 5 detailing a case study for an enterprise energy management SoS. Section 6 concludes and discusses future work.

## 2 SoS Interoperability

Within many modern information systems, interoperability is not seen as a strong requirement within their design. This results in limited interoperability between systems and a high cost associated with system alignment. The classic interoperability challenge can be seen within existing enterprises systems. An enterprise can use multiple information systems to support its business processes including sales, marketing, management reporting, research and development, financial reporting, customer loyalty and satisfaction, and similar areas. Integrating these systems is a key challenge for any IT department as the number of systems involved can be significant, and it can be time-consuming and expensive to stitch together the legacy systems so they use common data and processes.

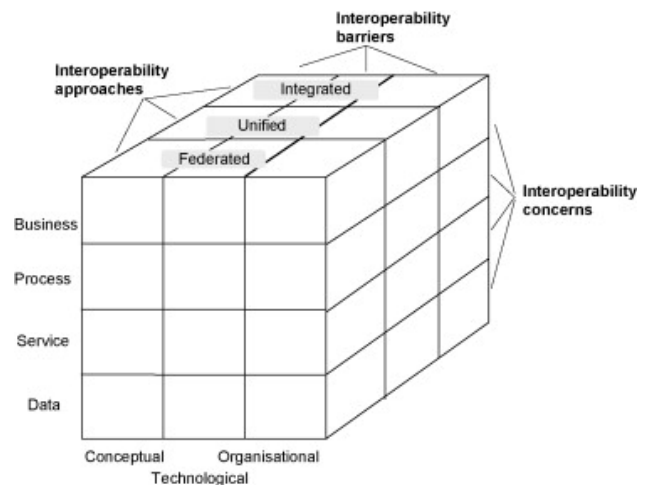


Figure 1. Enterprise interoperability framework [1]

Interoperability is a complex issue with many facets. The INTEROP framework [1], see Figure 1, identifies the barriers and approaches to interoperability. INTEROP identifies three dimensions of interoperability barriers:

- 1) *Conceptual barriers* represent the syntactic (format) and semantic differences (interpretation of meaning) of exchanged information.

- 2) *Technological barriers* refer to the incompatibility of information technologies such as protocols, encoding, platforms, or infrastructures.
- 3) *Organizational barriers* concern the organizational incompatibility of definitions of responsibility, authority, and organizational structures.

Interoperability approaches are the means and solutions to address interoperability barriers. INTEROP classifies interoperability approaches into:

- *Integrated approaches* where a common format for all models exists.
- *Unified approaches* where an upper model exists to facilitate mapping between models.
- *Federated approaches* where no common format exists but rather is done on the fly.

## 2.1 Types of System of Systems

Many definitions of a SoS exist, within this work we use the definition of a SoS as a collaborative set of systems in which its components are independent dedicated systems that are separately acquired and integrated to form a single system, yet maintain a continuous operational existence independent of the collaborative system [2]. A SoS can take many forms, SoS classifications [3] include *Virtual*, *Collaborative*, *Acknowledged*, and *Directed*.

- *Directed SoS* are built and centrally managed to fulfil specific purposes. Component systems maintain an ability to operate independently, but their normal operational mode is subordinated to the central managed purpose.
- *Acknowledged SoS* have defined objectives and dedicated resources. The constituent systems retain their independent ownership and objectives. Changes in the systems are based on collaboration between the SoS and the systems.
- *Collaborative SoS* component systems interact voluntarily to fulfil agreed upon central purposes. The central players collectively decide the means of enforcing and maintaining standards.
- *Virtual SoS* have no central management authority and a centrally agreed upon purpose. Virtual SoS rely upon relatively invisible mechanisms to maintain them.

## 2.2 SoS Interoperability Challenges

The differing SoS types present different interoperability requirements. On the one hand a Directed SoS with dedicated resources and central coordination can implement an Integrated or Unified approach to interoperability (i.e. Many-to-One). At the other extreme a Virtual SoS has no central authority or resources and may require a Federated approach to interoperability (i.e. Many-to-Many), in such scenarios a constituent system may not even be aware they are involved in a SoS.

The existing system design mind-set views interoperability as an external responsibility, and outsources interoperability to external systems. Within SoS, interoperability needs to be a fundamental requirement to their design and operation. The key challenge is to simplify interoperability without increasing complexity, hierarchy, control, or acquisition cost. This will require a change in mind-set of system design to embrace interoperability concerns. An effective interoperability approach for SoS will minimize complexity and ensure constituent systems do not need to be re-engineered as other constituent systems are added, removed, modified, or replaced. To this end, a common interoperability infrastructure is needed to support flexible SoS information interoperability. The infrastructure needs to be a lightweight technology platform supporting the minimum set of requirements to enable the maximum flexibility of the interoperability approach. In this initial work we focus on the foundation level of the INTEROP framework and investigate information interoperability for SoS. However, improving the conceptual and technical interoperability of systems is an important step to support organizational interoperability.

## 3 Background

### 3.1 Dataspace

A Dataspace is an emerging data management architecture that is very distinct from current approaches. The dataspace approach recognizes 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 [4]. Dataspace shifts the emphasis to providing support for the co-existence of heterogeneous data that does not require an upfront investment into a unifying schema. Data is integrated on an “*as needed*” basis with the labor-intensive aspects of data integration postponed until they are required. Dataspaces reduce the initial effort required to setup data integration by relying on existing matching and mapping generation techniques. This can result in a loosely integrated set of data sources, when tighter integration is required it can be achieved in an incremental “*pay-as-you-go*” fashion by more closely integrating the required data sources. Similarly a dataspace may only provide weak guarantees of consistency and durability. As stronger guarantees are desired, more effort can be put into making agreements among the various systems, see Table 1.

Table 1. Dataspace comparison

	<b>Model</b>	<b>Formats</b>	<b>Control</b>	<b>Query</b>	<b>Integration</b>
DBMS	Relational	Homog.	Complete	Precise	Explicit
Dataspace	All	Heterog.	Partial	Approx.	Implicit/ Incremental

### 3.2 Linked Data

Information integration projects typically focus on one-off point-to-point integration solutions between two or more systems in a customized but inflexible and ultimately non-reusable manner. The fundamental concept of Linked Data is that data is created with the mind-set of sharing and reuse. Emerging from research into the Semantic Web, Linked data leverages the existing open protocols and W3C standards of the Web architecture for sharing structured data on the web. Linked Data proposes an approach for information interoperability based on the creation of a global information space [5]. The main components are 1) *Universal Resource Identifiers* (URIs) to name things, 2) *Resource Description Framework* (RDF) for representing data, 3) *Linked Data principles* for publishing, linking, and integration, 4) *Vocabularies* to establish and share understanding, and 5) *Bottom-up incremental agreement*. Linked Data has the following advantages:

- Separate systems that are designed independently can be later joined/linked at the edges.
- Interoperability is added incrementally when needed and where it is cost-effective.
- Data is expressed in a mixture of vocabularies.

### 3.3 Linked Open Data

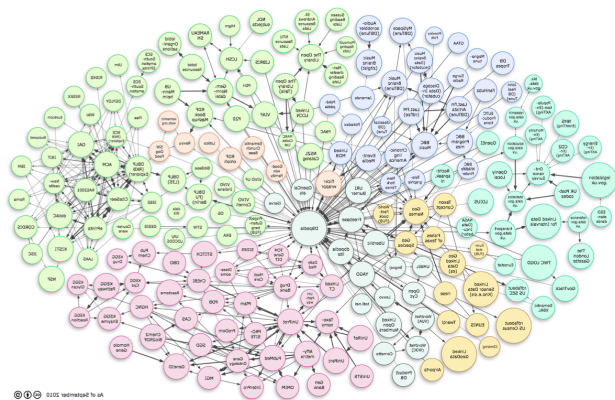


Figure 2. The linked open data cloud

Linked Data is facilitating the publishing of large amounts of structured data on the web. The resulting Web of Data can be considered as a web scale dataspace supported by Semantic Web technologies. The Linked Open Data Cloud, see Figure 2, represents a large number of interlinked datasets that are being actively used by industry, government and scientific communities [6]. The cloud has been growing considerably; as of September 2011 it has over 300 open data sets with more than 35 billion facts, interlinked by 500 million links.

### 3.4 Linked Data Principles

Linked data technology uses web standards in conjunction with four basic principles for exposing, sharing and connecting data. These principles are:

**1. Use URIs as names for things:** the use of Uniform Resource Identifier (URI) (similar to URLs) to identify things such as a person, a place, a product, an organization, an event or even concepts such as risk exposure or net profit, simplifies reuse and the integration of data.

**2. Use HTTP URIs so that people can look up those names:** URIs are used to retrieve data about objects using standard web protocols. For an employee this could be their organization and job classification, for an event this may be its location time and attendance, for a product this may be its specification, availability, price, etc.

**3. When someone looks up a URI, provide useful information using the standards:** when someone looks up (dereferences) a URI to retrieve data, they are provided with information using a standardized format. Ideally in Semantic Web standards such as RDF.

**4. Including links to other URIs so that people can discover more things:** retrieved data may link to other data sources, thus creating a data network e.g., data about a product may link to all the components it is made of, which may link to their supplier.

### 3.5 Resource Description Framework

The Resource Description Framework (RDF) is the basic machine-readable representational format used to represent information. RDF is a general method for encoding graph-based data that does not follow a predictable structure. RDF is schema-less and self-describing, meaning that the labels of the graph describe the data itself. Data and facts are specified as statements and are expressed as atomic constructs of a subject, predicate and object, also known as a triple. The statement “Edward Curry is the Occupant of Room 202e” is expressed in triple format as:

Subject - “Edward Curry”  
Predicate - “is the Occupant of”  
Object - “Room 202e”

RDF is designed for use in web-scale decentralized graph data models. For this reason the statement parts need to be identified so that they can be readily and easily reused. RDF uses URIs for identification, expressing the previous statement in RDF then becomes:

[http://www.deri.ie/about/team/member/edward\\_curry#me](http://www.deri.ie/about/team/member/edward_curry#me)  
<http://vocab.deri.ie/rooms#occupant>  
<http://lab.linkeddata.deri.ie/2010/deri-rooms#r202e>

## 4 Linked Dataspace for SoS

This paper proposes that Dataspaces together with Linked Data can be the basis for a SoS interoperability approach that would help to overcome technical and conceptual barriers to information interoperability.

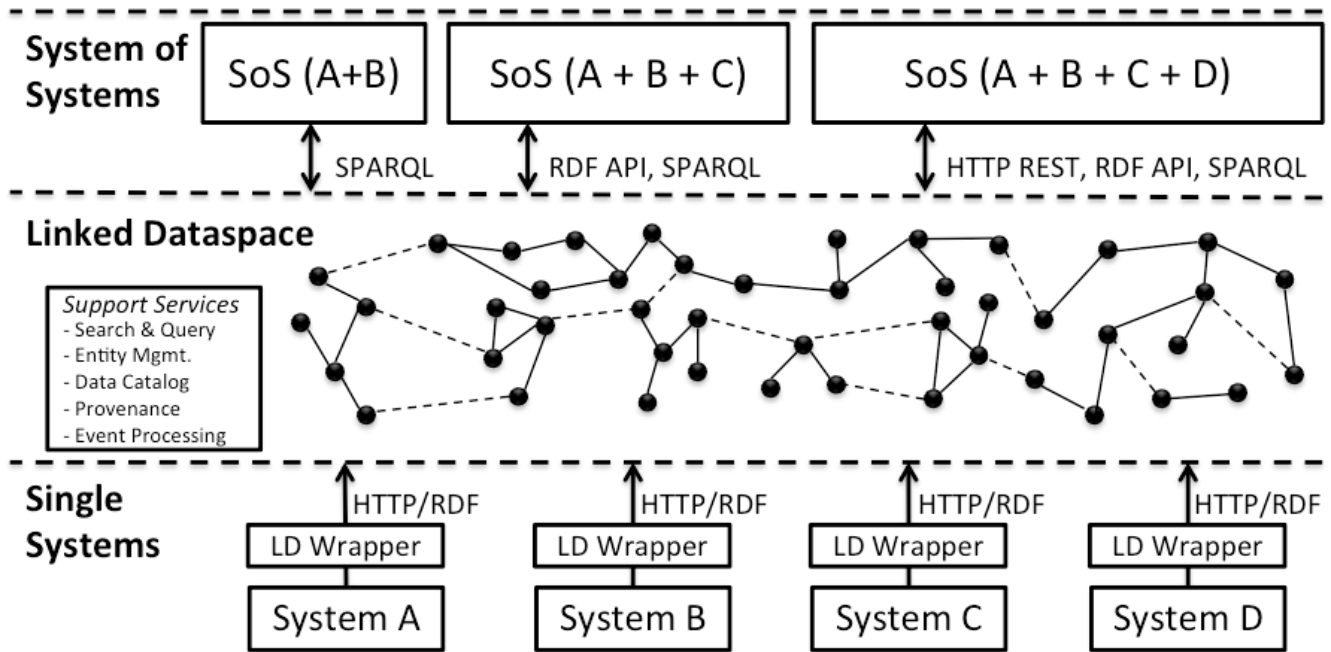


Figure 3. Linked dataspace for system of systems

A linked dataspace can be accommodated with minimal disruption to existing information infrastructure, as a complimentary technology for data sharing, it should not be seen as a replacement for current infrastructure (relational databases, data warehouses, etc.). The objective is to expose the data within existing systems, but only link the data when its needs to be shared. Once a system exposes its data within the linked dataspace it can participate in many SoS, knowingly or unknowingly

#### 4.1 Architecture

Figure 3 shows the proposed placement of a linked dataspace layer for interoperability of SoS. The linked dataspace serves as an independent layer placed above the existing systems layer, but below the system of systems layers. The approach can support the three approaches to interoperability (Integrated, Unified, Federated) with linked data providing a common syntactic and access protocols. The main components of the architecture are the wrapper on existing systems, the Linked dataspace consisting of a linked data cloud & support services, and the resulting system of systems.

#### 4.2 Linked Data Wrappers

At the bottom of the architecture are the existing operational legacy information systems. Wrappers perform the “RDFization” process, which transforms multiple formats and legacy data and lifts it to the dataspace.

#### 4.3 Linked Data Cloud

The Linked Dataspace links at the information-level (data) not the infrastructure-level (system) by focusing

more on the conceptual similarities (shared understanding) between information. This is achieved by following an entity-centric approach that focuses on the concepts that exist within the systems, for example, business entities like employees, products, customers, intellectual property, assets, etc. Entities within the dataspace are enriched with data from multiple systems. This results in a cloud of interlinked resources that reflect virtual or actual entities with links to relevant knowledge and contextual information from across all the information systems that have exposed linked data. Sources may be added in an incremental manner to the cloud where they can be reused. Each entity within the cloud has a dereferenceable URI that returns data in a machine-readable format describing the resource identified.

Agreement on data schemas can be achieved in a number of ways, from minimal agreement on a small number of simple vocabularies, to the use of agreed upon domain ontologies. The most appropriate approach will depend on the type of SoS involved and the interoperability approach undertaken. In keeping with the minimal agreement ethos of a dataspace, and the triple structure of RDF, the Entity- Attribute-Value (EAV) model is appropriate for implementing an entity-centric dataspace [7]. EAV can facilitate a *pay-as-you-go* methodology for data integration allowing the incremental enrichment of entity profiles. Multiple triples can be joined together to build up an entity-centric graph of information. Figure 4 provides an example of a entity-centric linked data graph segment, the example shows how relevant energy data can be exposed and linked to build up the energy profile of a user from across corporate, building and office IT systems.

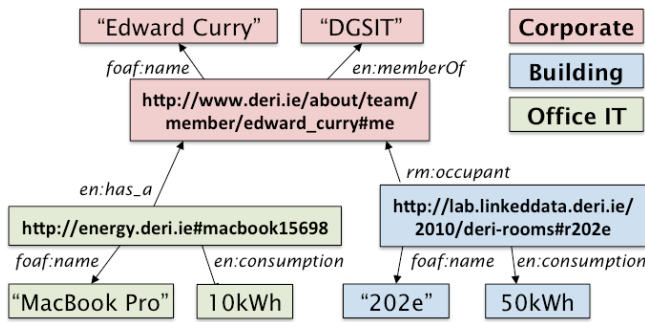


Figure 4. Entity-centric linked data graph

#### 4.4 Support Services

Dataspace support services are designed to simplify the consumption of the linked data cloud by encapsulating common services for reuse. Some example support services used in this work are:

- *Entity Management Service* to improve data quality and inter-linkage between entity data scattered among legacy systems. The EMS can leverage automatic entity consolidation algorithms that are supported by humans for collaborative data management [8].
- *Complex Event Processing* engine [9] is used to assess situations of interest that are encoded as event/action rules. Real-time information from sensors networks are also supported via the Semantic Sensor Network Ontology [10].
- *Data Catalogue and Provenance service* [11] to query the catalogue about data sources with specific attributes such as freshness and publisher, and track data back to its origin.
- *Search and Query services* [12] allow users to interact with the dataspace using structured or natural language interfaces.

#### 4.5 System of Systems

At the top of the architecture is the SoS that consume the resulting data and events from linked data cloud. A SoS developed using the Linked Dataspace is now discussed.

### 5 Enterprise Energy Management

An enterprise can use energy in many ways, from manufacturing and logistics, to building and IT. Within each of these areas an enterprise may have information systems deployed to help manage energy usage. However, in order to manage energy holistically, a real-time energy view of the entire enterprise is needed to provide an ability to understand, change, and reinvent business processes to better support energy efficient practices. This requires a SoS with support for a cross-domain view that helps users to interpret the information and understand how it can be used to optimize energy usage.

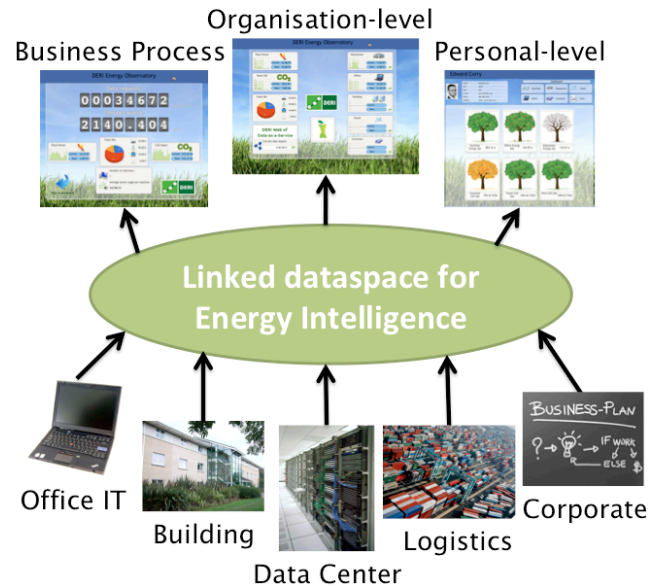


Figure 5. Enterprise energy system of systems

To demonstrate a linked dataspace for SoS interoperability we have implemented an enterprise energy management SoS within the Digital Enterprise Research Institute (DERI). The resulting Linked dataspace for Energy Intelligence (LEI), see Figure 5, enables a holistic SoS view of the energy management within DERI (facilities, office IT, Data Centre, business travel, commute, etc.), which can be used to manage energy at different-levels (organization, business process, individual) and from different domains perspectives (energy, finance, sustainability, human resources, etc). The LEI dataspace connected over 10 existing systems and was used to deliver a number of other energy management tools including buildings, office IT, data centre, and cloud services. In all cases the Linked Dataspace provided the linkages required to integrate operational legacy information system, and real-time data streams in a straightforward and flexible manner.

### 6 Related Work

The Global Information Grid (GIG) represents a globally interconnected, end-to-end set of information capabilities and processes for collecting, processing, and managing information on demand. The GIG follows a network-centric approach to data sharing, replacing the “one-to-one,” interoperability approach (protocols, standards, etc.) with a “many-to-one” interoperability approach in which each system will interface with the GIG. The Net-Centric Adapter for Legacy Systems (NCALS) [13] is designed to enable net-centric operations, it focuses on enabling legacy systems to interoperate with the GIG. NCALS provides a service-oriented architecture connection for a legacy system to the GIG. However, it does not modify the legacy components to comply internally with the net-centric technical standards. It allows the legacy system architecture to remain largely undisturbed.

The Global Earth Observation System of Systems (GEOSS) [14] seeks to be a global public infrastructure that generates comprehensive, environmental data, information, and analyses for a wide range of users. The GEOPortal disseminates information and analyses directly to users, making it easier to integrate diverse data sets, identify relevant data and portals of contributing systems, and access models and other decision-support tools. Access to data is through service interfaces with interoperability specifications agreed among contributing systems.

Both the GIG and GEOSS are large-scale SoS taking different implementation approaches. We believe a linked dataspace can have a role for data interoperability within such systems and is particularly suited where the web architecture is already a part of the SoS. Within GEOSS, EuroGEOSS is currently investigating RDF as a basis for semantic interoperability.

## 7 Summary

Within system of systems significant technical challenges exist in terms of information interoperability that require overcoming conceptual (syntax and semantics) and technological barriers. This paper presented an approach to System of Systems information interoperability based on the dataspace approach to information management and linked data approach to sharing information on the web. The resulting linked dataspace supports the co-existence of heterogeneous data with an incremental approach to interoperability. The paper describes the fundamentals of the approach and demonstrates the concept with a system of systems deployment for Enterprise Energy Management.

Future work will focus on the development of appropriate support services for linked dataspace using approximation techniques [15], [16], and the interoperability of system services as RESTful services.

**Acknowledgements:** This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

## References

- [1] D. Chen, G. Doumeingts, and F. Vernadat, "Architectures for enterprise integration and interoperability: Past, present and future," *Computers in Industry*, vol. 59, no. 7, pp. 647-659, 2008.
- [2] E. Rechtin, "The art of systems architecting," *IEEE Spectrum*, vol. 29, no. 10, pp. 66-69, 1992.
- [3] M. W. Maier, "Architecting Principles for Systems-of-Systems," *Systems Engineering*, vol. 1, no. 4, pp. 267-284, 1998.
- [4] M. Franklin, A. Halevy, and D. Maier, "From databases to dataspace: a new abstraction for information

management," *Sigmod Record*, vol. 34, no. 4, pp. 27-33, 2005.

- [5] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, no. 1, pp. 1-136, Feb. 2011.

- [6] E. Curry, A. Freitas, and S. O'Riain, "The Role of Community-Driven Data Curation for Enterprises," in *Linking Enterprise Data*, D. Wood, Ed. Boston, MA: Springer US, 2010, pp. 25-47.

- [7] P. M. Nadkarni, L. Marengo, R. Chen, E. Skoufos, G. Shepherd, and P. Miller, "Organization of heterogeneous scientific data using the EAV/CR representation.," *Journal of the American Medical Informatics Association*, vol. 6, no. 6, pp. 478-93, 1999.

- [8] U. ul Hassan, S. O'Riain, and E. Curry, "Leveraging Matching Dependencies for Guided User Feedback in Linked Data Applications," in *9th International Workshop on Information Integration on the Web (IIWeb2012)*, 2012.

- [9] S. Hasan, E. Curry, M. Banduk, and S. O'Riain, "Toward Situation Awareness for the Semantic Sensor Web: Complex Event Processing with Dynamic Linked Data Enrichment," in *4th International Workshop on Semantic Sensor Networks 2011 (SSN11)*, 2011, pp. 69-81.

- [10] L. Lefort, C. Henson, K. Taylor, and P. Barnaghi, "Semantic Sensor Network XG Final Report, W3C Incubator Group Report (28 June 2011)," 2011.

- [11] A. Freitas, T. Knap, S. O'Riain, and E. Curry, "W3P: Building an OPM based provenance model for the Web," *Future Generation Computer Systems*, vol. 27, no. 6, pp. 766-774, Jun. 2011.

- [12] A. Freitas, J. Oliveira, S. O'Riain, E. Curry, and J. Pereira da Silva, "Querying Linked Data using Semantic Relatedness: A Vocabulary Independent Approach," in *Proceedings of the 16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011*, 2011, vol. 6716, pp. 40-51.

- [13] A. Thomas, T. Turner, and S. Soderlund, "Net-Centric Adapter for Legacy Systems," *IEEE Systems Journal*, vol. 3, no. 3, pp. 1-7, 2009.

- [14] M. L. Butterfield, J. S. Pearlman, and S. C. Vickroy, "A System-of-Systems Engineering GEOSS: Architectural Approach," *IEEE Systems Journal*, vol. 2, no. 3, pp. 321-332, 2008.

- [15] S. Hasan, S. O'Riain, and E. Curry, "Approximate Semantic Matching of Heterogeneous Events," in *6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012)*, 2012.

- [16] A. Freitas, E. Curry, J. G. Oliveira, and S. O'Riain, "Querying Heterogeneous Datasets on the Linked Data Web: Challenges, Approaches, and Trends," *IEEE Internet Computing*, vol. 16, no. 1, pp. 24-33, 2012.