Coordinating Social Care and Healthcare using Semantic Web Technologies

Spyros Kotoulas, Vanessa Lopez, Martin Stephenson, Pierpaolo Tommasi,
Wei Jia Shen, Gang Hu, Marco Luca Sbodio, Veli Bicer, Anastasios Kementsietsidis,
M. Mustafa Rafique, Jason Ellis, Thomas Erickson, Kavitha Srinivas,
Kevin McAuliffe, Guo Tong Xie, and Pol Mac Aonghusa

IBM Research

1 Introduction

Healthcare and Social Care are unique domains in terms of cultural importance, economic magnitude and complexity. On a cultural level, the level of advancement of a society is often measured in terms of protection of the less able. In economic terms, for 2009, total expenditure on healthcare in the United States was 2.6 trillion USD or 17.4% of the GDP[1]. Total expenditure on social care was 2.98 trillion USD or 19.90% of the GDP[2]. In terms of US Federal government expenditure, social security, medicare and medicaid amount to 45% of total spending. In terms of complexity, organizations that are involved in providing social and medical care are numerous and span a very wide domain. For example, AHIP, the trade association of health insurers numbers some 1300 members[3]; the number of hospitals registered with the American Hospital Association is 5724[4] and the number of homeless shelters surpasses 4000[5]. In addition, medical information is vastly complex: Nuance reports that LinkBase®[6] contains more than 1 million concepts. Social care depends on information from a very broad domain, ranging from criminal records to housing.

Coordinating social care and health care has been identified both as a major pain point and a significant opportunity in modern health and social systems [1]. Several studies have shown that costs can be contained and outcomes improved with a more holistic approach to care [2]. As a simple motivating example, consider an individual quartered in inappropriate housing while suffering from a relatively minor health issue, aggravated by the housing condition. As a result, the given individual frequently resorts to visiting emergency rooms, resulting in significant cost to the healthcare system and a less effective treatment. By itself, the housing situation does not warrant state intervention. Nevertheless, resolving it would dramatically improve the health situation, resulting in a better quality-of-life for the individual and lower costs for the health system.

1 http://dx.doi.org/10.1787/888932523215
2 http://www.oecd.org/els/social/expenditure
3 http://www.ahip.org
4 http://www.aha.org/research/rc/stat-studies/fast-facts.shtml, retrieved 19/04/2013
5 http://www.shelterlistings.org/
6 http://www.nuance.com/for-healthcare/resources/clinical-language-understanding/ontology/index.htm

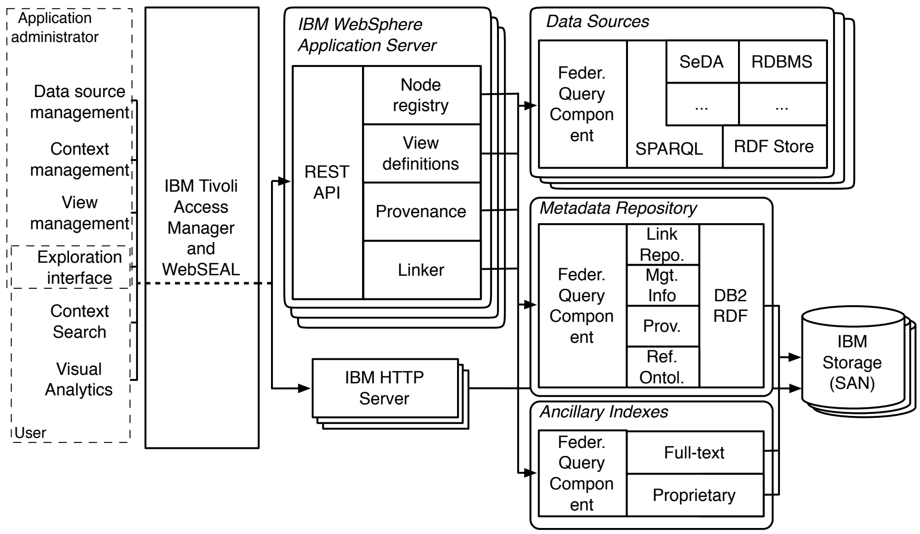

Fig. 1: System architecture

Even in this simple example, the challenges presented are significant: *How do we access information in disparate systems, storing vastly heterogeneous information on various infrastructures? How do we cope with policy constraints disallowing replication or centralization of data? How do we abstract from the information and representation complexity?*

In this paper, we propose a novel technical solution to augment applications with cross-domain context, in the domain of Social Care and Healthcare based on business rules and contextual exploration. We claim that Semantic Technologies can uniquely address these problems because: (a) The distributed nature of RDF allows access to integrated information across silos. (b) Explicit and global semantics allow us to ground business rules across systems. (c) The distributed and incremental data integration paradigm advocated by linked data can help coping with the complexity of the data.

We present a demonstrator of a system that supports two key use-cases for this domain: (a) Displaying a view of the combined needs across several dimensions for a given person and people in their social context, based on a set of business rules. This allows a social/health worker to quickly assess the situation of an individual. From a knowledge management perspective, it requires grounding a set of business rules across several ontologies and instance data in several data sources. (b) Exploration of the context to surface information not directly covered by the business rules. Given the heterogeneity of the domain, the user will most likely need additional information around a given individual. Our demonstrator uses the business rules as a navigational aid to explore the semi-structured information.

2 Approach

Key Performance Indicators (KPIs) are used to ground business rules to data, offering a tree-based view over the factors that contributed to a given KPIs and the weight of each factor (influence) to the global vulnerability score, helping us understand relations across different needs. For example, being homeless (or living under poor housing conditions) is an aggravating factor for health. These views can be applied to an individual, family members, or socially/geographically organized groups. Each node in a KPI is associated to two SPARQL queries. The first one is to calculate the score of a given contributing factor (if present) obtained from a given data source(s) and with a given weight. The second one is a CONSTRUCT query to retrieve the set of triples providing additional context in the ontology(-ies) associated to the data that contributed to the score, as well as the justification on the values from which this KPI factor was derived. The score of a KPI node is the sum of its own score and that of its children. For both types of queries, rather than being tied to a specific model, we abstract from the particular representation using a set of query patterns.

An enterprise architecture supporting our approach is shown in Fig. 1. Due to space restrictions, we describe only the components necessary to understand the basic operation of the system. Web-facing services use a set of REST services, implemented on a custom application running on IBM WebSphere Application Server. The main components for these services are the *Node registry*, which tracks nodes in the *Federated Query Engine*, the *View definitions*, that are used to project information out of the graph model for use by analytics widgets and UI elements. *Data Sources* are exposed as virtual RDF, using SeDA, an IBM technology to execute R2RML mappings. The virtual RDF Data Sources, the Metadata Repository and the Ancillary Indexes are accessed through the *Federated Query Engine*, providing transparent access to the distributed information. All core components in this architecture can be clustered, for high availability and performance.

3 Deployment

We have internally deployed a proof of concept based on the above architecture, integrating a set of IBM solutions for clinical and social program information: **IBM software Patient Care and Insights** provides data driven population analysis to support patient centered care processes. It integrates and analyzes the full breadth of patient information sourced from multiple systems and different care providers. It stores three categories of data: extracted patient medical history called clinical summary; medical data analytics results from an analytics component called care insights and personalized electronic care plans. **IBM Cúram** is a business and technology solution to help social program organizations provide optimal outcomes for citizens, satisfy increasing demand, and lower costs for organizations. In connection to this paper, the information of interest mainly regards social relationships, known problems concerning employment, substance abuse, participation in social assistance programs and information concerning housing, education and safety.

Figure 2 shows some UI components from our proof of concept. Since our approach is meant to be deployed as part of a existing application, in order to augment them with information from other systems, we have opted to focus on the context that can be retrieved, rather than trying to replicate the enterprise application: (a) *Genogram, Fig. 2, floating frame on top-left.* We have adapted the genogram visualization [3] to

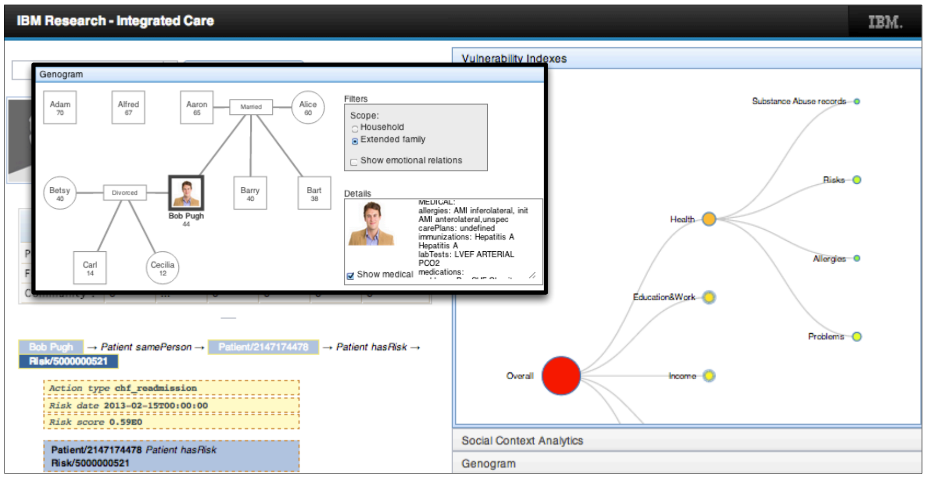

Fig. 2: Screenshot

explore the family environment of a person and associated problems. (b) *Hierarchical KPI, Fig. 2, right.* The tree allows the user to explore the vulnerabilities of a person using information coming from several sources. The KPIs themselves are tree structures. Clicking on a node brings up a contextual exploration view. (c) *Contextual exploration, Fig. 2, bottom-left.* The user is able to investigate information related to a node in the KPI tree based on a graph exploration interface. In addition to elements shown in the figure, our proof of concept supports exploration and analysis based on the spatial component and family relations.

From internal feedback, the main strong points of our approach lie in the ability to consume data from heterogeneous sources without complicated data warehouses, associated ETL processes and setting up related infrastructures, although it remains to be seen whether the tooling required for a semantic approach will reach the sophistication of what is currently found in the enterprise domain. In addition, tabular or tree-like visualizations are strongly preferred to graphs. Future work lies in better-informed data exploration, mining the RDF graphs to identify meaningful relationships and data inconsistency checking across silos.

References

1. Rigby, M., Hill, P., Koch, S., Keeling, D.: Social care informatics as an essential part of holistic health care: A call for action. I. J. Medical Informatics **80**(8) (2011) 544–554
2. Peikes, D., Chen, A., Schore, J., Brown, R.: Effects of care coordination on hospitalization, quality of care, and health care expenditures among medicare beneficiaries. JAMA: the journal of the American Medical Association **301**(6) (2009) 603–618
3. Jolly, W., Froom, J., Rosen, M., et al.: The genogram. The Journal of family practice **10**(2) (1980) 251