

Guest Editorial Preface

Special Issue on Semantic Web Technologies for Government

Branco Ponomarev, The University of Texas at San Antonio, San Antonio, TX, USA

The current interest in the constantly improving capacity for relating and analyzing large quantities of data (often branded as “Big Data”), in the business and government bureaucracies presents an interesting paradox. Of course bureaucracies need data, however the efficient functioning of a bureaucracy is predicated on reducing, rather than increasing the volume and types of data that need processing in the course of organization’s routines. To the extent information is required in establishing bureaucratic processes, such requirements are essential in the analysis and design stage of these processes (as earnestly articulated about a century ago by the spokesman of ‘scientific management’, F. Taylor). The paradox is further compounded by the old truism that information and data are not “knowledge” or “insight”, and thus have a limited value of their own.

If so, what are the rationales behind the enthusiasm for “Big Data”, “Linked data” and related trends in IT? To be sure, the benefits and the promise are substantial. Most notably, these include the ease and low cost of collection (often simply as a byproduct of ongoing transactions), and the ability to capture and mine a great variety of measures. Yet the most pressing problems for most organizations are not related to data deficit, but rather to “strategy” or “vision” deficit. Why the enthusiasm then?

In the public sector, there are at least two drivers for such an interest. First, government are under increasing pressure to make more and different types of data available for public use. However the usability of data generated and provided by government agencies often lags behind availability: even when data is made available, it is often not “raw data”, or even if it is, it is often not presented in machine readable format, and its consistency, reliability, and context are not easily evaluated, especially by users without substantial background knowledge and expertise in the respective policy area. Consequently, one impetus behind the interest in semantic web technologies is rendering heterogeneous government data – originating in diverse professional contexts and collected for different purposes - more easy to identify, integrate and use, including eventually by means of agents or bots capable of recognizing relevant data and undertaking some analysis and decision tasks on behalf of humans. Open data initiatives such as data.gov in the US attempt to address such difficulties by establishing data depositories and standard rules for describing, authenticating and distributing data.

Second, and at the present still somewhat less urgent, semantic web technologies have substantial promise to alleviate common problems with the application of IT in decision making or process design, namely domain uncertainty, integration and inter-operability of data and processes across jurisdictions, and even simply a relevant process or service discovery. The suite of technologies underpinning the notion of the ‘semantic web’ are expected to facilitate or even automate some of the contextual decision making humans engage in when analyzing data by enabling software to recognize “the meaning” of different sets of data and to make ‘decisions’ about their connections and relevance. Although fifteen years after the visionary article in “Scientific American” by Berners-Lee the implementation of this idea has been limited, a number of advances have been made.

In short, big data is not about “data” alone – big data does not become “Big Data” until large volumes of data are linked together algorithmically. Only when such successful combinations of algorithms and data occur we can talk about qualitatively new data applications in support of government functions. The semantic web technologies represent one such approach towards integration, and the articles presented in this special issue provide an insight in the state of the art of the semantic web thinking and implementation in governmental contexts, including application and implementation case studies. The opening article by Bulazel and colleagues tackles a problem that has been frustrating analysts and developers for decades: inconsistencies in naming schemes, or how to reliably and authoritatively to identify the same entities across different data sets when they are defined by and referred to by different vocabularies. Indeed, these are among the fundamental issues the semantic web is hoped to address. Their proposed solution simultaneously alleviates some of the common problems characterizing the information on the World Wide Web. Their “instance hub”, or an authoritative entity reference for entities commonly used by agencies and in open data initiatives tackles the vagueness, uncertainty, the inconsistency, and the untrustworthiness that commonly characterizes even simple data sets, e.g. ones describing jurisdictions or clearly defined objects.

Thomas and Alalwan, in their paper “Designing a Semantic Tool to Evaluate Web Content of Government Websites” tackle another common problem seemingly affecting even the most comprehensive open data initiatives: the poor management of data changes over time, often resulting in deteriorating currency, consistency, and ultimately relevance and usability. They develop and test a content management ontology incorporating existing content management ontologies and manuals, and discuss the challenges to create and inter-operate such ontologies and their wider application.

Although all contributors to the special issue acknowledge that the design and broad application of semantic web technologies is not a purely technical issue, and is in fact inconceivable without explicitly considering the role of managerial, decision making, and organizational factors in general, Alexandrova and colleagues provide the most in-depth assessment of the organizational issue involved in their report on the implementation of an ontology-based Citizen Relationship Management system at multi-jurisdictional call centre at the Miami-Dade county, including descriptions of the varied engineering tasks and user roles. Their report highlights the relative ease and simplicity of applying (free, open-source and standards-based) core semantic web technologies to modeling complex process interactions across agencies and jurisdictions, the crucial role of business domain experts in developing the ontologies, the importance of training business users in the system to ensure continuity and consistency, as well as the inherent tension between “technically correct” modifications of the ontology in response to emerging business requirements vis a vis expedient or emergency modifications, and accordingly offer potential remedies and recommendations. (They also highlight the capacity of open source tools to circumvent common problem with commercial IT solutions offered to governments: services and maintenance arrangements designed with an eye to maximize billable hours, rather than to offer flexible platform which governments could modify and maintain in-house - problems avoidable with the in-house design approach they present.)

The paper by Wout Hoffman proposes entrusting data (though not process) management to end users even further, by developing a governance model transferring data stewardship roles onto the citizens and enterprises, while government retains its data inspection and decision making roles, and

the role of data custodian is played by a different, private or public, party. His model illustrates well the attractiveness and the feasibility of defining exhaustively and semantically codifying the events, activities, and subjects involved in government transactions. Although the conceptualization of such models is often criticized as “vaporware”, by definition their potential future implementation is inconceivable without systematic efforts to articulate their structure and implications, including analysis of practical cases, as in this contribution.

Puron-Cid and colleagues’ review and re-examine the promise of open data and policy informatics for public policy analysis and decision making. Their expose reminds us that while the developments in technology, as usual, are “promising”, such promise does not inherently translate into government service or capacity improvement in the absence of cultivating human capital and institutional changes to apply such advancements. In particular, the plethora of new tools and technologies cannot by themselves “solve” problems such as who are the relevant communities and participants in the process, what is the “proper” definition of the policy problem, etc.

The issue concludes with a contribution by Sayogo and colleagues, discussing ontological modelling of certification and inspection processes (in support of disclosure of product information). The paper proceeds from a real policy problem (the difficulties associated with private regulation of defining and certifying product standards, especially in international trade) as its starting point. The proposed solution to reconciling the lack of clarity in the meaning of voluntary certification and labeling systems, and the extent of compliance with them, is an ontological modelling process of certification and inspection schemes to enable the integration and standardization of certification and inspection data. Although the authors acknowledge this effort stops short of identifying and planning the necessary governance mechanisms that would be need for the implementation of such ontology-based certification and inspection data architecture, the contribution documents a valuable blueprint – based on thorough field research of relevant actors and stakeholders - to inform future policy in the area of private regulation of third party certification systems.

As usual, focus on the technical means of rationalizing policy decision making makes it easy to overstate the extent to which the policy process itself is actually rational, however one of the promises of semantic web technologies in particular is to at least partially address the human capital shortages: while the analysis of large and disparate data sets and processes inherently requires both conceptual and practical knowledge of different analysis tools and data structures, semantically defined and machine readable data sets and processes potentially enable even users without sophisticated technical background to successfully apply and modify such tools as needed (as also highlighted in the Alexandrova and colleagues’ contribution) and successfully query and model data from disparate information depositories for policy analysis purposes.

The promise of new open data and semantic web applications will not by itself make policy making more rational, but it certainly can, and has, affected the distributions of power and influence in policy controversies. In particular, advances in open data, policy informatics, and semantic modelling tend to empower incumbent organizational structures (e.g. government agencies and corporations), while their impact on civil society and civic participation, is not straightforward.

The papers presented in this special issue highlight this discrepancy, by locating the power for such redesigns firmly in the hands of technocrats – developer and analysts communities - with the partial exception of the Hoffman’s contribution: even then the citizen role is limited to data provision and stewardship – a role that is not necessarily empowering, considering it may represent shifting some of the agency workload onto citizens, while retaining power on the uses and the meaning of the data. Similarly, the potential of social media applications to solicit and aggregate citizen input also does not necessarily mean more democratic policy making, unless citizens also have the opportunity to frame the terms of the debate. While policy informatics is a powerful tool to address well defined and structured problems, the problem definition itself is inherently political and contested arena and even choice of technologies itself has political implications, as Puron-Cid and colleagues remind us.

Indeed, the common conceptualization of issues, problems, and processes is the fundamental assumption without which creating (widely used and accepted) semantic web ontologies is virtually impossible. As such, common conceptualization is central to both the political and the technical aspects of the nascent semantic web experiments. Although the semantic web is based on a “open world” assumption, how far such common conceptualizations can be taken is an interesting question likely to evolve differently in different policy arenas and communities.

The contributions in this special issue highlight the complex interplay of technology, politics, and civil society, raise both big questions and provide useful practical recommendations. They illustrate that at present, much progress has been made in the development and applications of semantic web technologies to improve government services and transparency, yet much of the promise of the semantic web is for the time being not yet fully realized. We hope that this special issue will inspire further thinking and research on this important topic.

Branco Ponomariov

Guest Editor

IJPADA