



8th Extended Semantic Web Conference

May 29 - June 2 2011 | Heraklion, Greece

Workshop

Semantics in Governance and Policy Modelling



May 30th, 2011, Heraklion, Greece

Agenda – Proceedings

With the support of:



**CROSSROAD – A Participative Roadmap for ICT Research
in Electronic Governance and Policy Modelling**



**GIC – Greek Interoperability Centre based at the
Decision Support Systems Lab of the National
Technical University of Athens**

Workshop on «Semantics in Governance and Policy Modelling»

Yannis Charalabidis, University of Aegean, yannisx@aegean.gr

Fenareti Lampathaki, National Technical University of Athens, flamp@epu.ntua.gr

Gianluca Misuraca, JRC - European Commission, IPTS, Spain, Gianluca.MISURACA@ec.europa.eu

Preface

Today, Governments struggle to regulate an increasingly interconnected, complex, fast-evolving and unpredictable world, as the recent financial crisis has shown. At the same time, the amount of data available to any interested stakeholder for making sense of the socio-economic environment has increased exponentially, provided either through sensors in the Internet of Objects and “crowdsourced” citizens’ reports in Social Media, or with the help of open government initiatives that have already started to emerge and impact governments’ operation, efficiency and reputation.

Moving in the strategic directions of the “Research Roadmap in ICT for Governance and Policy Modelling” prepared by the CROSSROAD Project (<http://www.crossroad-eu.net>), data-powered collective intelligence and action has been unanimously recognized as a key challenge towards a better, evidence-based, timely, truly transparent and participative decision making process. In such a Collaborative Governance vision, citizens are engaged in all the phases of policy making: in collecting available data through open linked data and participatory sensing techniques; in analyzing the data through data mining, information visualization and visual analytics tools; and in acting on relevant data upon conducting interactive simulations and serious games experimentations that accompany and stimulate government policies.

Semantics are bound to play an important role in these new governance grand challenges, supporting policy modeling, citizen opinion mining, data visualization, automated and personalized service provision.

Topics and Papers

The papers presented in this workshop are targeting novel approaches, methods, practices and tools for turning the public information overload into an opportunity for Governance and Policy Modelling with the help of semantics, open data and Public Sector Information reuse, linked data and visual analytics. Specific importance is placed in multidisciplinary approaches, combining theoretical policy making methodologies, advanced ICT and Government 2.0 collaboration practices.

The workshop provided a stand for real-life application examples, showing how stakeholders can gain insight and take action into complex and unpredictable real-life situations at international, national or local level, including: Semantic components for digital public service definition and provision; Semantics for policy modelling, including policy planning and policy impact assessment; Open, governmental data publication and promotion; Linking of open governmental data; Linked data in government data mash-ups; Collaborative analysis and collective intelligence for policy-making, Semantic annotation of public policy and citizen opinion; Semantic components for public deliberation and argument visualization.

Agenda

The papers selected among submissions and made publishable in the workshop are following, organized under two sessions, as for presentation. The first session gathers papers in the areas of Open and Linked Data in Governance, including also structural components approaches:

- Fenareti Lampathaki, Yannis Charalabidis, Dimitris Askounis: *"Open Linked Data and Visual Analytics in Electronic Governance: State of Play and Perspectives"*
- Konstantin Hypponen, Miika Alonen, Sami Korhonen, and Virpi Hotti: *"XHTML with RDFa as a Semantic Document Format for CCTS Modelled Documents and its Application for Social Services"*
- Kelli de Faria Cordeiro, Maria Luiza Machado Campos, Marcos R. S. Borges: *"Empowering Citizens and Government with Collaboration on Linked Open Data"*
- Lucas de Ramos Araújo, Pablo N. Mendes, Jairo Francisco de Souza: *"Publishing Linked Data from Brazilian Politicians on the Web"*

The second part touches upon issues relative to more participative approaches, mainly at the policy support and deliberation areas. As such, Semantically-enriched Policy Making is a new direction for several existing categories of research artefacts, attracting papers such as:

- Riccardo Boero, Enrico Ferro, Michele Osella, Yannis Charalabidis, Euripidis Loukis: *"Policy Intelligence in the Era of Social Computing: Towards a Cross-Policy Decision Support System"*
- Christos Tsarouchis, Declan O'Sullivan, David Lewis: *"Balancing system expressivity and user cognitive load in semantically enhanced policy modeling"*
- George Anadiotis, Panos Alexopoulos, Konstantinos Kafentzis, Efrosyni Konstantinou, Milan Stankovic, and Manuel Waeckerle: *"Semantics-powered Virtual Communities and Open Innovation for a Structured Deliberation Process"*
- Alessio Gugliotta, Franco Alberti, Igor Jogan: *"Comelicopedia: a Collaborative, Semantically-enriched Knowledge Base for the Sustainable Development in Mountain Areas"*

Audience, Organizers

The audience of the workshop includes researchers and practitioners spanning the domains of Semantic Web, Electronic Governance and Policy Modelling, university students and professors of Computer, Management, Social and Public Administration sciences, ICT industry staff engaged in relevant projects and solutions, policy makers and decision drivers at local, national or international levels, public administration officials and industry executives, engaged both in organisational design and ICT deployment.

Workshop Chairs

Yannis Charalabidis (Corresponding Chair), University of Aegean, yannisx@aegean.gr

Fenareti Lampathaki, National Technical University of Athens, flamp@epu.ntua.gr

Gianluca Misuraca, JRC - European Commission, IPTS, Spain, Gianluca.MISURACA@ec.europa.eu

Workshop Programme Committee

Dimitris Askounis, National Technical University of Athens, Greece

Enrico Ferro, ISMB, Italy

Marko Grobelnik, Jožef Stefan Institute, Slovenia

Euripidis Loukis, University of Aegean, Greece

Michele Missikoff, IASI-CNR, Italy

David Osimo, Tech4i2, Belgium

Maria Wimmer, University Koblenz-Landau, Germany

Table of Contents

Overview	6
1. Open Linked Data and Visual Analytics in Electronic Governance: State of Play and Perspectives	7
2. XHTML with RDFa as a Semantic Document Format for CCTS Modelled Documents and its Application for Social Services	20
3. Empowering Citizens and Government with Collaboration on Linked Open Data	33
4. Publishing Linked Data from Brazilian Politicians on the Web.....	38
5. Policy Intelligence in the Era of Social Computing: Towards a Cross-Policy Decision Support System.....	51
6. Balancing system expressivity and user cognitive load in semantically enhanced policy modelling.....	64
7. Semantics-powered Virtual Communities and Open Innovation for a Structured Deliberation Process.....	77
8. Comelicopedia: a Collaborative, Semantically-enriched Knowledge Base for the Sustainable Development in Mountain Areas	90
Authors Index.....	95

Overview

Open Linked Data and Visual Analytics in Electronic Governance: State of Play and Perspectives

By Fenareti Lampathaki, Yannis Charalabidis, Dimitris Askounis

XHTML with RDFa as a Semantic Document Format for CCTS Modelled Documents and its Application for Social Services

By Konstantin Hypponen, Miika Alonen, Sami Korhonen, and Virpi Hotti

Empowering Citizens and Government with Collaboration on Linked Open Data

By Kelli de Faria Cordeiro, Maria Luiza Machado Campos, Marcos R. S. Borges

Publishing Linked Data from Brazilian Politicians on the Web

By Lucas de Ramos Araújo, Pablo N. Mendes, Jairo Francisco de Souza

Policy Intelligence in the Era of Social Computing: Towards a Cross-Policy Decision Support System

By Riccardo Boero, Enrico Ferro, Michele Osella, Yannis Charalabidis, Euripidis Loukis

Balancing system expressivity and user cognitive load in semantically enhanced policy modelling

By Christos Tsarouchis, Declan O'Sullivan, David Lewis

Semantics-powered Virtual Communities and Open Innovation for a Structured Deliberation Process

By George Anadiotis, Panos Alexopoulos, Konstantinos Kafentzis, Efrosyni Konstantinou, Milan Stankovic, and Manuel Waeckerle

Comelicopedia: a Collaborative, Semantically-enriched Knowledge Base for the Sustainable Development in Mountain Areas

By Alessio Gugliotta, Franco Alberti, Igor Jogan

1. Open Linked Data and Visual Analytics in Electronic Governance: State of Play and Perspectives

Open Linked Data and Visual Analytics in Electronic Governance: State of Play and Perspectives

Fenareti Lampathaki¹, Yannis Charalabidis², Dimitris Askounis¹

¹ National Technical University of Athens, 9 Iroon Polytechniou str., 15780 Athens, Greece
{flamp, askous}@epu.ntua.gr

² University of the Aegean, Karlovassi, 83200 Samos, Greece
yannisx@aegean.gr

Abstract. In a world confronted with global societal challenges and aligned with disruptive technological trends, ICT for Governance and Policy Modelling emerges as a key enabler for managing complexity and driving state reorganisation, openness and effectiveness in collaboration with citizens, businesses and society in general. In the years to come, citizens are expected to get engaged in all the phases of policy making: in collecting available data through open linked data and participatory sensing techniques; in analyzing the data through data mining, information visualization and visual analytics tools; and in acting on relevant data upon conducting interactive simulations and serious games experimentations that accompany and stimulate government policies. This paper investigates the state of play in Open Data, Linked Data and Visual Analytics in order to provide research, policy and practice insights with regard to ICT for Governance and Policy Modelling. It identifies targeted research issues that challenge each area according to an hierarchical taxonomy, while analyzing their emerging perspectives.

Keywords: ICT for Governance and Policy Modelling, Open Data, Linked Data, Visual Analytics, State of the Art, Taxonomy.

1 Introduction

In an era where there is a need for extreme transparency, with fading boundaries between countries and between ICT and social behaviour, plus enhanced intelligence gathering [21], there is a greater need than ever to ensure that policies are designed in an integrated manner across different governance levels and between policy areas to maximize beneficial synergies and minimize negative trade-offs. Today as the previous goals of “making services available online” have started to fail, the new strategic direction in Electronic Governance appears to be transparency and “making public data available for reuse” [20].

In an increasingly interconnected, complex, and fast-evolving society, the amount of data available for making sense of the socio-economic environment is increasing exponentially, provided through sensors and mobile devices, with the help of open government initiatives that have already started to emerge and release public data, or

even in bottom-up ways through “crowdsourced” citizens’ reports in Social Media. However, governments clearly still struggle to make sense of such large amounts of data and take appropriate action.

Future scenarios in ICT for Governance and Policy Modelling are promising to reach the target of a better, participative, evidence-based and timely governance, while taming greater complexity and attracting citizens’ involvement [5]. ICT for Governance and Policy Modelling has emerged as an umbrella term for a number of technologies that can be applied in order to achieve the common goal of improving public decision-making in the age of complexity [7].

In this context, this paper presents the state of the art related to the “Open Government Information & Intelligence for Transparency” Research Theme within the taxonomy of ICT for Governance and Policy Modelling proposed by the CROSSROAD project [8]. It tries to incorporate next generation semantic technologies into the Governance and Policy Modelling context by promoting the principles of open data and PSI reuse, as well as the philosophy of linked data and visual analytics.

The structure of this paper is as follows: Section 2 outlines the research areas taxonomy in ICT for Governance and Policy Modelling. Section 3 gives an overview of the implications and challenges in providing public data «free-as-in-speech». Section 4 identifies the current landscape in Linked Data in the public sector. Section 5 finally presents the emerging Visual Analytics that aim to turn the underlying information overload into an opportunity, leading to conclusions with the perspectives opened up for future research in the domain.

2 Research Areas Taxonomy in ICT for Governance and Policy Modelling

In order to bring together the different perspectives related to ICT for Governance and Policy Modelling, a taxonomy that maps the domain and identifies the issues that fall within its scope has been constructed on the basis of a set of metrics and thresholds. With the help of a quantitative and qualitative methodology, a Research Areas Taxonomy has been based on the material collected and elaborated in the state of the art analysis in an effort to reach consensus on the diverse domain of ICT for Governance and Policy Modelling [17], [18]. This taxonomy classifies the research areas at three levels based on the desk-research and the extensive discussions with experts, as well as the interactions with involved scientific communities in their domain of expertise.

The taxonomy now consists of 5 Research Themes (RTs) as broad thematic categories that contain and classify a number of research areas at lower levels (i.e. 17 Research Areas and more than 80 Research Sub-areas), as following:

The *Research Theme RT.1: Open Government Information & Intelligence for Transparency* appears as the data- and knowledge- oriented research theme. It infuses light-weight semantic technologies into the Governance and Policy Modelling context based on the philosophy of open data, as well as the foundations of linked data and visual analytics.

The *Research Theme RT.2: Social Networks, Citizen Engagement and Inclusion* infuses the social dimension of the web into Governance and Policy Modelling by investigating the Social Computing phenomenon that has already revolutionized the way people communicate, exchange content and knowledge, raise their opinions and influence each other, by exploiting engagement and eParticipation tools and techniques, and by extracting people's opinion from the web in order to reach collective wisdom.

The *Research Theme RT.3: Policy Making*, clearly positioned towards ICT-enabled policy making, initially analyzes the economical, social and environmental context as a preparatory stage for policy modelling which then actually represents the problem and its proposed solution. Policy Simulation testing out the various models in an effort to pre-evaluate the application of a specific policy, in a controlled, artificial environment is also studied, while at the last stage in the policy-making process, Policy Evaluation provides the necessary qualitative and quantitative assessment mechanisms to monitor the actual policy application.

The *Research Theme RT.4: Identity Management and Trust in Governance* is driven by the need to safeguard citizens' and public authorities' digital presence from misuse. In this context, identity management with federated identities, access control and authentication mechanisms in ubiquitous environments, as well as privacy and data protection have proved to contribute in building trust among citizens and public authorities.

The *Research Theme RT.5: Future Internet for Collaborative Governance* embraces the internet evolution and entails transparent and multichannel service provision via the Internet of Services, low cost cloud infrastructures emerging from Cloud Computing advancements, better human-computer interfaces and seamless interaction with non-conventional web devices that communicate in the Internet of Things.

3 Open Data and Transparent Information Management

The public sector collects, produces, reproduces and disseminates a wide range of information in many areas of activity, such as social, economic, geographical, weather, tourist, business, patent and educational information, commonly known as Public Sector Information (PSI) [10]. Traditionally, PSI has been published in different ways and formats, from the early paper days to the early Web days in which information was being published online in whatever format was more convenient for the government organization in charge of publishing it and according to the normative at that time [26]. Recently, though, the realization that the power of data comes from the ability to sort, search and transform it for a wide range of applications has started to dominate leading to the so-called open data initiatives.

Sharing data actually enables greater transparency, delivers more efficient public services, and encourages greater public, private and non-profit / civic use and re-use of government information. Open data holds up the promise of instant connectivity between arbitrary numbers of ad hoc partners while at the same time reducing

integration costs, harnessing external innovation, and even creating entirely new and significant business models, governance models and engagement models.

Motivated by underlying "open" movements and communities such as open source and open access, the term "Open Data" itself has recently been established as a philosophy and practice requiring that certain data are freely available to everyone, without restrictions from copyright, patents or other mechanisms of control in a timely and accessible way. Such open data needs to be "free as in speech" (with few or no restrictions, libre, open source) in contrast to "free as in beer" (for zero price, gratis, freeware) [27].

In this direction, Government data shall be considered open if the data are made public in a way that complies with the principles below which are further aligned with the Open Knowledge Definition [22]: 1. Data Must Be Complete, 2. Data Must Be Primary, 3. Data Must Be Timely, 4. Data Must Be Accessible, 5. Data Must Be Machine processable, 6. Access Must Be Non-Discriminatory, 7. Data Formats Must Be Non-Proprietary, 8. Data Must Be License-free.

As legal and fiscal constraints guide the use of technology in government, new questions arise, such as: Does open data make government vulnerable? Is it responsible to share data without appropriate context? What data should be openly shared? Such questions have already started to trouble government authorities. While several of the open data initiatives could be covered by simple attribution licences, the possible negative economic effects of attaching "copyleft" or "sharealike" licences to (or allowing such licences to infect) public data remain to be studied.

The following table provides an overview of the research sub-areas and questions that challenge the Open Data research area according to the respective taxonomy.

Table 1. Open Data and Transparent Information Management Taxonomy Extract

Research Sub-area	Definition
<i>1.1.1 Open Data Publication</i>	Mechanisms and tools for opening up, preparing, publishing and documenting open data available to the public in easily accessible formats and in a timeframe that allows their full economic potential to be exploited
<i>1.1.2 Web Dissemination and Promotion</i>	Methodologies and tools for widely disseminating the open data released to the public, also through web 2.0 channels
<i>1.1.3 Open Data Quality Agreements</i>	Methodologies and tools for creating on-the-fly data quality agreements following the paradigm of service quality agreements
<i>1.1.4 Open Data Communities Building</i>	Tools and techniques for creating, hosting and sustaining vibrant user communities and ecosystems that request, provide and use open data
<i>1.1.5 Transparency and Reputation Management</i>	Methods and techniques for automatic evaluation and monitoring of transparency, reputation and overall impact achieved via unleashing public data
<i>1.1.6 Open Data Legal Implications and Licences</i>	Approaches and methods for automatically managing and resolving legal implications of publishing data, i.e. right to publish, licence schemes, etc.

In terms of policy directions, the European Union Directive 2003/98/EC, commonly called the Public Sector Information Directive, has introduced a common

legislative framework regulating how public sector bodies should make their information available for re-use in order to remove barriers such as discriminatory practices, monopoly markets and a lack of transparency. Open data initiatives are further promoted in the Visby Declaration, in the Granada Strategy, the Ministerial Declaration on eGovernment approved unanimously in Malmö, Sweden, on 18 November 2009, the Digital Agenda and the European eGovernment Action Plan 2011-2015.

At a country-level, many Member States have already started to follow the PSI Directive and US paradigms in opening up data. For example, in the UK “Putting the Frontline First: Action Plan”, a key recommendation is to radically open up data and public information in order to promote transparent and effective government and social innovation. In US, the President Barack Obama has taken bold steps to inaugurate an era of government openness and transparency by issuing a Memorandum on Transparency and Open Government, the Open Government Directive and the OSTP Open Government Plan articulating the indispensable role that science, technology, and innovation (ST&I) must play if the great challenges of our time are to be met.

2009 actually proved to be the year for putting government data online [1] in “free as in speech” terms since indicatively the following initiatives appeared:

- At country level: US (data.gov) according to which “*Data accessed through Data.gov do not, and should not, include controls over its end use*”; UK (data.gov.uk) with terms of use interoperable with any Creative Commons Attribution 3.0 Licence; Australia (data.australia.gov.au) in which “*Datasets made available through data.australia.gov.au are subject to certain licensing conditions. The details of each licence can be found in the Dataset Information section of each dataset*”; New Zealand (data.govt.nz) with a Creative Commons Attribution 3.0 New Zealand licence.
- At state level in the US: California (ca.gov/data) where “*information presented, unless otherwise indicated, is considered in the public domain. It may be distributed or copied as permitted by law*”; District of Columbia (data.octo.dc.gov) with terms of use including that “*Applications using data supplied by here must perform the following: 1) Notify the District of Columbia via email; 2) Include the following disclaimers on their sites...*”; Utah (utah.gov/data) in which “*The public is granted access to information on the State's computer system on a strictly "as is" basis*”
- At city level in the EU: London (data.london.gov.uk) in which “*The license conditions under which data can be used will vary for each data package but in general the data held in this site can be used for most purposes, provided the Terms and Conditions of this site are not infringed*” and Lichfield (lichfielddc.gov.uk/data), while in the US San Francisco (datasf.org) use of data is subject to specific terms of use.

If the government can provide well-organized data in common formats, then technology and services can be built on top of the data, such as: iLive.at, a winner in the 2008 Apps for Democracy contest run by the Washington D.C.; FollowTheMoney.eu aiming to make it easier for European citizens to understand the EU budget: how it gets decided, where the money comes from and how it is spent; UK FixMyStreet.com where UK residents can report problems in their neighborhood

(like graffiti or potholes on the road); and Everyblock (US that combines metropolitan data like crime, road construction, building permits, and sanitation violations statistics and makes information easily accessible to the public for consumption).

4 Linked Data

Traditionally, data published on the Web have been made available as raw files in formats such as CSV or XML, or marked-up text as HTML tables, sacrificing much of their structure and semantics. However, in recent years the Web has evolved from a global information space of linked documents (“Web of Documents”) to one where both documents and data are linked (“Web of Data”). Underpinning this evolution is a set of best practices for exposing, sharing, and connecting structured pieces of data, information, and knowledge on the Web known as Linked Data.

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods [19]. It refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets [3].

In the early days of linked data (2006-2008) the main focus of the linked data community was on publishing data and finding good practices [2]. Since then, the Web of Data has opened up new possibilities and fertile research ground for Capturing, Sharing, Querying, Browsing and Mashing-up Linked Data, as depicted in the following table.

Table 2. Linked Data Taxonomy Extract

Research Sub-area	Definition
<i>1.2.1 Capturing and Sharing Linked Data</i>	Modelling techniques for making government data linkable and for representing and semantically annotating linked data in order to be easily consumed in ubiquitous computing environments. Tools for publishing large data sources using light weight ontologies on the Web and embedding linked data with semantics into classic Web documents. Business Models for sharing and publishing linked data.
<i>1.2.2 Automated Querying and Reasoning over Governmental Linked Data</i>	Approaches for querying distributed data over different Web sources regardless the “physical position” of data and immediately processing results and for applying sophisticated automated reasoning to machine understandable data, including automated deduction, rule induction, abduction, probabilistic reasoning.
<i>1.2.3 Browsing and Searching Linked Data</i>	Tools and techniques for seamlessly crawling, navigating and retrieving different data sources.
<i>1.2.4 Government Data Fusion and Mash-ups</i>	Methods and frameworks for automatic generation of links, intelligent mappings and data mash-ups aggregating similar types of media and information from multiple sources into a single representation.

Research Sub-area	Definition
1.2.5 <i>Linked Data Provenance and Evolution</i>	Methodologies and tools for tracking original data sources and for maintaining, updating and synchronizing governmental links and datasets
1.2.6 <i>User Interaction and Linked Data Usability</i>	User interface/interaction and data usability issues. Understanding why and how people use and information.
1.2.7 <i>Linked Data Quality Assurance</i>	Methodologies and tools for evaluating and assuring the quality, trustworthiness, and appropriateness of Web data

Since significant volumes of Linked Data have already been published on the Web, numerous approaches have been established in order to research and build applications that exploit this Web of Data, such as: DBpedia [4]; Revyu [12], and Riese (RDFising and Interlinking the Eurostat Data Set Effort) [11].

At a policy-making level, most governmental announcements and guidelines are associated with the open data initiatives and do not explicitly emphasize on linked data. The only exception up to now appears to be the UK government which has already declared its commitment to publishing data as linked data [25].

In the last couple of years, a lot of significant data mash-up applications have emerged (often through relevant innovation contests and awards) demonstrating how useful tools can result from giving citizens greater access to government data, i.e. Data-gov Wiki Demo: Castnet Ozone Map (US)¹; UK Research Funding Explorer (UK)²; and EnAKTing UK Public Sector Information Data Mashup (UK)³.

5 Visual Analytics

Taking into account that today, data is rapidly created and published at incredible volume and the ability to collect and store the data is increasing at a faster rate than the ability to analyze it, the need to go beyond mere Information Visualisation for representing data, information or knowledge in a way that will facilitate and enhance understanding has emerged. In this context, Visual Analytics is characterized as an emerging area of research and practice that aims at integrating the outstanding capabilities of humans in terms of visual information exploration and the enormous processing power of computers to form a powerful knowledge discovery environment, allowing them to make well-informed decisions in complex situations [13].

Visual Analytics is defined as the science of analytical reasoning supported by interactive visual interfaces [24], while according to Wikipedia, it refers to "*the integration of interactive visualization with analysis techniques to answer a growing range of questions in science, business, and analysis*". The basic idea of Visual Analytics is to visually represent the information, allowing the human to directly interact with the data to gain insight, see things they were not aware of, define new

¹ <http://data-gov.tw.rpi.edu/demo/exhibit/demo-8-castnet.php>

² <http://bis.clients.talis.com>

³ <http://map.psi.enacting.org>

questions, hypotheses, and models of their data, draw conclusions, and ultimately make better decisions [13].

Visual Analytics is an inherently multi-disciplinary field [16], including the following focus areas [24]:

- Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning, and decision making
- Visual representations and interaction techniques that take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once
- Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis
- Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences.

In this context, the following table provides an overview of the research sub-areas and questions that challenge the Visual Analytics research area according to the respective taxonomy.

Table 3. Visual Analytics Taxonomy Extract

Research Sub-area	Definition
<i>1.3.1 Visual Information Foraging and Design</i>	Theories and applications trying to automatically understand how human users search for information and how visualization storyboards and scenarios should be designed in order to cover their needs
<i>1.3.2 Information Visualization and Interaction</i>	Visual representations and interaction techniques including the principles for depicting information, new visual paradigms, statistical graphics, geospatial visualizations, aesthetics and animation, the science of interaction, and approaches for generating visualizations and interactions in multiple user interfaces and devices
<i>1.3.3 Analytical Reasoning</i>	Analytical reasoning techniques and tools via which users obtain deep, high-quality insights that directly support situation assessment, planning, and decision making with a limited investment of the analysts' time
<i>1.3.4 Collaborative Analysis and Intelligence</i>	Approaches that enable collaborative analysis of visualizations, gaining common ground and building on each other's findings by studying the psychological aspects of humans processing. Awareness techniques, in which the interactions of one collaborator on a visualization are visible to other collaborators viewing the data items in their own visualizations or views of the data
<i>1.3.5 Visualization Evaluation</i>	Methodologies and tools to systematically evaluate the impact, worth, and significance of a visualization approach such as visualization quality, measurement of the effectiveness of the display, scalability, usability, assessment of its impact on decision-making, efficacy in conveying information, rhetorical effect etc.

Several researchers have contributed to creating a theoretical understanding of how individuals make use of information visualizations to gain insight into data and solve

problems [15]. Recent visualization systems that support asynchronous collaborative analysis indicatively include [14]: Many Eyes: Web-Based Visualization and Publishing, Wordle, Protovis, Swivel, Geovisual Analytics Visualization, and Wikimapia. Dedicated argument visualization tools are: Debategraph, Araucaria, Argunet.

In the US, the emphasis laid on strategic level at visual analytics was demonstrated in 2004 when the US Department of Homeland Security (DHS) chartered the National Visualization and Analytics Center (NVAC) with the goal of helping to counter future terrorist attacks in the United States and around the globe. A major objective for NVAC is to define a long-term research and development (R&D) agenda for visual analytics to address the most pressing needs in R&D to facilitate advanced analytical insight [24]. However, in policy and strategic documents, there is no explicit reference to the need for visual analytics or for exploiting visualization results in decision-making.

A significant number of visualization techniques exist today including methods for: Visualization of Quantitative Data; Animation, flow and dynamic representation; Spatial data, relationships and topologies; Dense, scalable pixel based visualizations, which are beneficial for the consumption of open linked data and policy making, by providing interconnected data visualization and exploration capabilities.

Indicative visualization examples on top of open governmental data include: Where does my money go? (UK)⁴ and Ease of Doing Business (World Bank Group)⁵.

With regard to policy modelling, a set of highlight visualization platforms includes:

- The IBM Splash: Smarter Planet Platform for Analysis and Simulation of Health research project [6]: It aims at building a framework that supports the integration of multiple existing models, simulations, and data that represent parts of the broader health ecosystem.
- The IBM City Forward⁶, which is a free, web-based platform that lets anyone interested to visualize and interact with city data while engaging with a community of people who are passionate about the future of our cities.

6 Conclusions & Perspectives

A paradigm shift in the nature and quality of policy-making is eventually happening based on three complementary disruptive techno-economic trends: the explosion in the quantity of data available; the proactive role assumed by citizens in contributing to the policy-making process, by authoring and elaborating the data increasingly made available; and the advanced capabilities, affordability and accuracy of simulation and modelling techniques. In this context, Open Government Information & Intelligence for Transparency has gained a lot of momentum in the last years as it summarizes the vision of governments for openness, accountability and collective intelligence, leveraging the underlying data plethora in order to make sense of hidden and unexpected ideas and take appropriate action.

⁴ <http://www.wheredoesmymoneygo.org/prototype>

⁵ <http://rru.worldbank.org/businessplanet>

⁶ <http://www.cityforward.org/>

Open Data and Transparent Information Management, although being firmly associated to the ICT for Governance and Policy Modelling, suggests at first sight that it is a practitioners' playground and that there are actually few research approaches, typically built on top of open data. It is well acknowledged that public sector information (PSI) may serve as a significant key ingredient in the policy making process for understanding the existing situation and feeding policy models, yet it is not efficiently exploited up to date. Upon investigating the state of the art, it appears that there are research questions in this context that mainly concern Timely Publication, Open Data Quality Agreements, Open Data Communities Building, and Transparency and Reputation Management. In order to achieve a Public Sector Information openness and reuse, strong political will together with changes in regulatory and legal frameworks are further required, strengthening the need for research in conjunction with Identity, Privacy and Trust in Governance.

As far as Linked Data is concerned, the scientific community is now paving the way to the Web integration, moving from the current Web of Documents – made by and for people – to the Web of Data – where machines can play a crucial role in knowledge management. Turning the Linked Data vision into reality still flourishes at research level without having met significant practice or market penetration so far, although standards and technologies have been already in place for many years. In general, access to high quality, timely, open linked data is central to facing the societal challenges, yet such data are not now seamlessly embedded in the policy models in an automatic way. Despite the current availability of open data, techniques and tools for curing, annotating, publishing open data, as well as for data provenance, are still lag behind the vision of the Web of Data and in most cases do not comply with the Linked Data principles. However, in order to achieve real-time interconnectedness of data from unlimited sources and devices for a common purpose, there is a great need for tracking original data sources and managing data and links evolution. Although linked data is very important to achieve data interconnection and information reuse, the current state of standards and tools require high technical and scientific skills, and thus can't be used by the average users in an out-of-the-box manner, or in a practical timeframe.

Finally, Visual Analytics is a young research field, which utilizes results from established areas such as information visualization and data mining and has seen unprecedented growth in its first 5 years of mainstream existence. Although state of the art visualization tools and prototypes are already in place in the market, there are also clear indications that there is a need to reconsider visualization ways and to think about new visualization paradigms, representation standards, interaction and sophisticated automated reasoning techniques in order to provide people with better and more effective ways to understand and analyze large datasets, while also enabling them to act upon their findings immediately, in real-time. Such techniques should automatically understand how human users search for information and how visualization story-boards and scenarios should be designed in order to cover their needs. Intuitive, easier visualization ways that bridge the gap between the research community's offerings and the stakeholders' needs in Governance and Policy Modelling also need to be specified in order to assess their efficacy in conveying information and their rhetorical effect on policy making. Although visual analytics may shed light on unexpected and hidden insights into policies, they do not invest on

collaboration at a satisfactory level, either directly exploiting the power of social media for seamlessly interacting with visualizations, or enabling collaborative analysis of visualizations, gaining common ground and building on each other's findings.

Along these lines of evolution, a set of more concrete research challenges have been already identified and validated through online and offline brainstorming and validation workshops [9]. The CROSSROAD Grand Challenge entitled "Data-powered collective intelligence and action", which aims at ensuring that increased transparency translates into actual more open and more effective policy-making, incorporates the strategic directions for enhanced data availability and better analytical capability.

Acknowledgements. This work has been created in the context of the EU-funded Support Action CROSSROAD (Contract No: FP7-ICT-248484). It has also been supported by the Greek Interoperability Centre activities (Contract No: FP7-ICT-204999).

References

1. Berners-Lee, T. (2009). Putting government data online. Retrieved: January 25, 2010 from <http://www.w3.org/DesignIssues/GovData.html>
2. Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008). Linked data on the web. In C. Bizer, T. Heath, K. Idehen, & T. Berners-Lee, WWW2008 Workshop on Linked Data on the Web. Beijing, China. Retrieved: January 25, 2010 from <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-369/paper00.pdf>
3. Bizer C., Heath T., & Berners-Lee T. (2009). Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), Special Issue on Linked Data, 5 (3), pp. 1-22
4. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., et al. (2009). DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 7(3), pp. 154-165.
5. Botterman, M., Millard, J., Horlings, E., van Oranje, C., van Deelen, M., Pedersen, K.: Value for citizens-A vision of public governance in 2020, European Commission (2008).
6. Cefkin, M., Flickner, M., Glissmann, S., Haas, P. J., Jalali, L., Maglio, P. P., Selinger, P., & Tan, W.-C. (2010). Splash: Smarter Planet Platform for Analysis and Simulation of Health. IBM Research – Almaden. Retrieved on January 12th, 2011 from <http://www.almaden.ibm.com/asr/projects/splash/splashvision20100728.pdf>.
7. Consultation report on ICT for Governance and Policy Modelling. FP7 Work Programme 2011-2012, February 2010, http://ec.europa.eu/information_society/activities/egovernment/research/fp7/docs/consultation_report_feb2010.pdf
8. CROSSROAD Project, <http://www.crossroad-eu.net> (2011).
9. CROSSROAD, D4.3 Final Roadmap, December 2010, http://crossroad.epu.ntua.gr/files/2010/02/CROSSROAD_D4.3_Final_Roadmap_Report-v1.00.pdf
10. European Commission (2003). EC Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-use of Public Sector Information.

Retrieved: February 23, 2010 from http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

11. Halb, W., Raimond, Y., & Hausenblas, M. (2008). Building Linked Data For Both Humans and Machines. In C. Bizer, T. Heath, K. Idehen, & T. Berners-Lee, WWW2008 Workshop on Linked Data on the Web. Beijing, China. Retrieved: November 11, 2010 from <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-369/paper06.pdf>.
12. Heath, T., & Motta, E. (2008). Revyu: Linking reviews and ratings into the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 6(4), pp. 266-273.
13. Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. Information Visualization, 7(1), pp. 49-62.
14. Heer, J., Wattenberg, M., & Agrawala, M. (2009). Point, Talk, Publish: Visualization and the Web. In E. Zudilova-Seinstra et al., Trends in Interactive Visualization. pp. 269-283.
15. Isenberg, P., Tang, A., & Carpendale, S. (2008). An exploratory study of visual information analysis. 26th Annual CHI Conference on Human factors in computing systems (CHI '08), 1217-1226. New York, New York, USA: ACM Press.
16. Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (Eds.): Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics Association (2010)
17. Lampathaki, F., Charalabidis, Y., Passas, S., Osimo, D., Wimmer, M., Askounis, D.: Defining a Research Areas Taxonomy around ICT for Governance and Policy Modelling. In Lecture Notes on Computer Science Volume 6228, pp. 61-72, Proceedings of the IFIP EGOV 2010 Conference, Lausanne Switzerland (2010)
18. Lampathaki, F., Koussouris, S., Charalabidis, Y., Askounis, D., Mouzakitis, S., Passas, S., Tsavdaris, H., Osimo, D., De Luca, A., Armenia, S., Bicking, M., Wimmer, M., Misuraca, G.: State of the Art in "ICT for Governance and Policy Modelling". In F. Lampathaki, S. Koussouris, Y. Charalabidis, & D. Askounis (Eds.), CROSSROAD White Paper (2010).
19. Linked Data Community (2010). Retrieved: February 7, 2010 from <http://linkeddata.org>
20. Meskill, D. (2009). Transparency and Open Government. Retrieved: February 15, 2010 from http://www.usaservices.gov/events_news/documents/Transparency_000.pdf.
21. Misuraca, G., Broster, D., Centeno, C., Punie, Y., Lampathaki, F., Charalabidis, Y., Askounis, D., Osimo, D., Szkuta, K., Bicking, M. (2010) Envisioning Digital Europe 2030: Scenarios for ICT in Future Governance and Policy Modelling. IPTS, JRC Scientific & Technical Reports.
22. Open Knowledge Definition (2010). Retrieved: February 15, 2010 from <http://www.opendefinition.org>
23. Paget (2010). Retrieved: February 7, 2010 from <http://code.google.com/p/paget>
24. Thomas, J., Cook, K. (2005). Illuminating the path: Research and Development Agenda for Visual Analytics. IEEE Press.
25. UK HM Government (2009). Putting the Frontline First: Action Plan. Retrieved: February 12, 2010 from <http://www.hmg.gov.uk/frontlinefirst/actionplan.aspx>
26. W3C Working Draft (2009) Publishing Open Government Data. Retrieved: February 10, 2010 from <http://www.w3.org/TR/gov-data>
27. Wershler-Henry, D. (2002) Free as in Speech and Beer: The Internet, Gift Economies and the Intellectual Property Crisis. Pearson Publications Company.

2. XHTML with RDFa as a Semantic Document Format for CCTS Modelled Documents and its Application for Social Services

XHTML with RDFa as a Semantic Document Format for CCTS Modelled Documents and its Application for Social Services

Konstantin Hyppönen, Miika Alonen, Sami Korhonen, and Virpi Hotti

University of Eastern Finland, School of Computing

P.O.B. 1627, FI-70211, Kuopio, Finland

Tel.: +358 17 162587; Fax: +358 17 162595

[Konstantin.Hypponen, Miika.Alonen, Sami.S.Korhonen, Virpi.Hotti]@uef.fi

Abstract. For achieving semantic interoperability, messages or documents exchanged electronically between systems are commonly modelled using standard specifications, such as the UN/CEFACT CCTS (core components technical specification). However, additional requirements, such as the need for layout markup or common metadata for certain archiving scenarios might be applied to the documents. Furthermore, the management of resulting artefacts, i.e., core components, XML schemas and related infrastructure, could be cumbersome in some cases. This paper investigates the use of the W3C XHTML+RDFa (extensible hypertext markup language with resource description framework attributes) for representing both the layout and semantics of documents modelled according to CCTS. The paper focuses on the validation of XHTML+RDFa documents against a core components library represented as an ontology. In addition, the paper illustrates and validates this demand-driven solution in the scope of the Finnish National Project for IT in Social Services.

1 Introduction

For a long time, documents have been the main information exchange mechanism in public administration. As the service processes get automated, documents become replaced or aided by electronic message exchange. The standard requirement for the electronic messages is that they must convey necessary semantic meaning of their contents in a form understandable by the information systems involved.

In many cases, however, additional requirements are applied to the message exchange system, such as the generation of human-readable versions of the messages, or archiving them according to a certain archiving plan. This might require the addition of layout markup and certain document management metadata to the messages.

Nešić [12] defines a semantic document as a “uniquely identified and semantically annotated composite resource”. The document is a content unit (CU) built of smaller CUs, which in turn can be either composite or atomic. Every

CU should be understandable by both humans and machines. Furthermore, the content of semantic documents should be completely queryable, with addressable elements (i.e., CUs) of different granularity.

Message exchange can be modelled using standardised specifications [13, 14, 18] which describe processes and models for designing common building blocks used in different messages. The specifications also include standard ways of designing XML schemas for business documents and information entities by reusing these building blocks [13, 17, 19]. The resulting schemas are purely semantic-oriented and are targeted specifically at message exchange between information systems. Therefore, they do not include provisions for any layout markup. Human-readable versions of the messages must be generated using separate style sheets or other mechanisms.

On the contrary, document formats used for preparing documents in common text processors (such as OOXML or ODF) or for publishing the documents online (such as HTML or PDF) concentrate mostly on layout features. They are not suitable as such for information exchange, as they lack precise semantic markup. However, semantic markup and metadata can be added to them using annotation tools, thus bringing semantics and layout markup together. Storing semantic annotations inside a document usually requires the extension of the document format schema, which is not always possible. Uren et al. [20] define requirements for annotation tools and provide a summarizing overview of them with regard to these requirements. Eriksson [4] describes a technology for ontology-based annotation of PDF documents. Individuals of ontology classes are created from highlighted areas of PDF documents. A bidirectional link is then established between the individual in an ontology and the PDF annotation.

Decker et al. [3] argue that flexible semantic interoperability cannot be achieved with XML, and suggest using RDF (resource description framework) based technologies for defining the semantics of information exchange. Combinations of XML and RDF technologies are also available. A knowledge model of the business domain could be represented in a form of ontology and linked to the schemas, for adding more precise semantics to schema elements. For example, the W3C recommendation SAWSDL (semantic annotations for the web services description language) defines a way of providing semantic annotations to WSDL and XML schema constructs [5]. Among other mechanisms, annotations can be implemented as links to ontology classes.

W3C provides a recommendation for a way to embed semantic markup in XML (e.g. XHTML) documents [1]. RDFa (RDF attributes) is a specification of attributes for adding semantic metadata to any markup language. RDFa can be used in conjunction with XHTML for extending the basic layout-oriented HTML markup with semantic elements. However, the applicability of XHTML and RDFa for message exchange designed with standard modelling methods such as the UN/CEFACT Core Components Technical Specification (CCTS) [18] has not been examined closely.

Our results. We propose the use of XHTML+RDFa as a document format (carrier) for messages modelled according to the CCTS specification. In addition,

we outline the structure of the validation service which ensures that the documents follow their defined structures, and performs other document type specific checks. A proof-of-concept implementation of the validation service is described. The applicability of this approach is examined in connection with requirements for the document exchange needed in Finnish social services.

The rest of the paper is structured as follows. Section 2 introduces the requirements for documents used in social services and provides an overview of the core components types defined in CCTS. Section 3 explains how core components can be represented in RDF, and Sect. 4 describes a validation service for XHTML+RDFa documents with CCTS-based artefacts. Section 5 discusses the pros and cons of our approach, and Sect. 6 provides concluding remarks.

2 Documents in social services

One of the goals of the IT project for social services in Finland (Tikesos¹) is to model all the document types that are used in Finnish social services. Document structures have been defined by the social service experts on the content level. The social services system in Finland uses about 200 different types of documents [11]. The documents are created in standard social work, where they are used as part of social service processes. As the processes get automated, the document contents are analysed and modelled for document processing systems. The documents can be used for information exchange between systems.

In addition to information exchange, one of the main requirements is to archive all documents for future use. This is demanded by the Finnish Archive Law [6], which is applied to all public authorities. A centralised archive is currently being developed for storing all social services documents in Finland [11]. When implemented, the archive will to be accessed by social services information systems, for archiving the documents (as shown in Fig. 1) and fetching them later if needed. We note that the archiving requirement is of utmost importance, as the future IT infrastructure for social services in Finland is built around the centralized archive.

There are no international de jure or de facto standards for a document format for social services. Therefore, we set out to define a document format suitable for our needs. The following basic requirements are placed on the document format for Finnish social services:

1. A nationally defined set of metadata must be implemented.
2. Content units of different granularity should be marked in the document, so that they are addressable.
3. It should be easy for an information system to parse the document contents.
4. A human-readable version of the document should be easily produced.
5. The validation of the document structure and content should be possible, according to validation service requirements.

¹ www.tikesos.fi

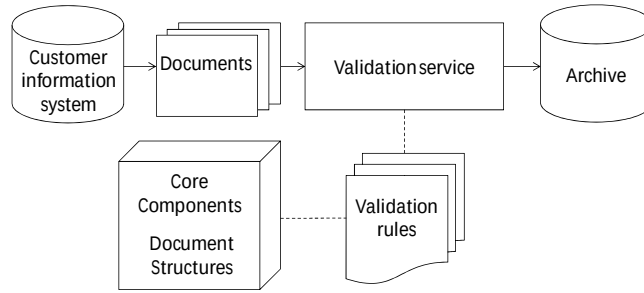


Fig. 1. The centralised archive validates and stores documents produced by customer information systems.

The common building blocks of the documents used in social services were identified, analysed and modelled according to the UN/CEFACT CCTS [18]. The CCTS model was selected because it is widely used in Europe for similar projects [8], and has the status of an ISO standard. We provide here a short introduction to the concepts defined in CCTS. Only the concepts relevant to this paper are described.

CCTS defines a method for designing of common semantic building blocks called core components. The components form a language used in information exchange by business partners. A number of different core component types (CCTS artefacts) is defined (see Fig. 2).

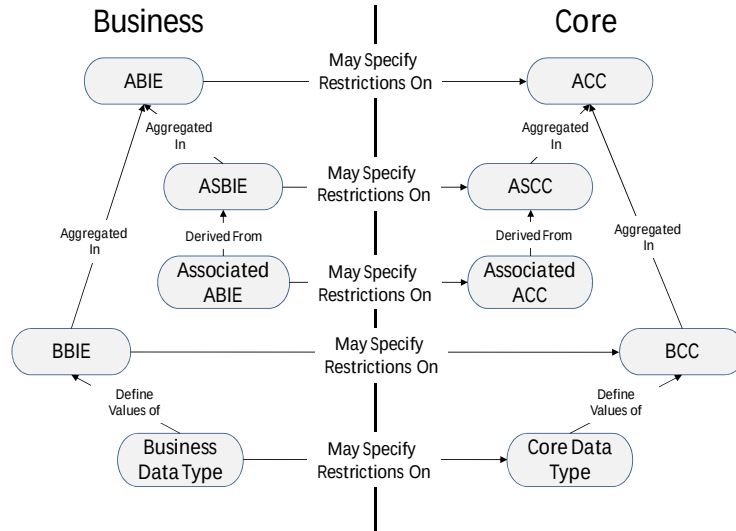


Fig. 2. Relationships of different types of core components (figure from [18]).

An Aggregate Core Component (ACC) is an object class describing a concept with clearly defined semantics. The object class contains a number of properties related to this concept. An example of an ACC is Address.

A Basic Core Component (BCC) is a property appearing in a certain object class (ACC). BCCs are based on simple data types (called Core Data Types) such as Text, Number, Amount. For example, the street name in an address can be represented as a BCC.

An Associated Core Component (ASCC) is a property appearing in a certain object class (ACC). The property is based on another object class (another ACC). For example, the validity period of an address could be represented as an ASCC, based on the ACC Period.

An Aggregate Business Information Entity (ABIE) is an object class which is a qualified version of an ACC. It may have a narrower meaning than the parent ACC. Moreover, any property of the ACC may be qualified in the ABIE. An example of an ABIE is Trade_ Address (an address representation used in trade-related business documents).

Basic Business Information Entities (BBIE) and Associated Business Information Entities (ASBIE) are properties of ABIE, constructed by qualifying the corresponding BCC and ASCC properties of the parent ACC.

The resulting library of core components defined for social services in Finland has around 200 different object classes (ACCs and ABIEs), with about 1000 properties in them altogether. The core components are used in some 200 different documents, split in 15 groups (applications, decisions, notifications, agreements etc.).

Although XML schemas for all business documents can be produced relatively easily, their applicability for our requirements is questionable. For instance, it is rather difficult to handle documents based on 200 different schemas in a single archive. The systems which access the archive must support these schemas in order to parse the document contents. Furthermore, document structures tend to change from time to time, and with the updated versions the number of different schemas can easily reach thousands in a span of a few decades. The documents based on old schemas remain in the archive and should be accessible by users, at least as human-readable versions. Producing human-readable document outputs can be also troublesome, as the layout information is not included in documents based on standard CCTS-compliant schemas, and separate schema-specific style sheets are needed. A separate human-readable output generator could be implemented within the archive. However, the generator must handle all the different schemas on which the documents in the archive are based.

3 From core components towards RDFa

CCTS does not define a way in which core components should be stored. UN/CEFACT distributes their international core components library in a table format, and provides specifications for the representation of business information entities as XML schemas. We present a transformation to an RDFS/OWL

representation from any core component library. The representation is aimed at simplifying the generation and validation of semantic XHTML+RDFa documents using existing core components. Business parties can model their documents according to the standard CCTS method, and use XHTML with RDF attributes as the data exchange medium.

The RDFS/OWL representation of a core component library is generated in a two-step process. First, a general XML serialisation of the library is produced, based on a custom schema. Second, the serialisation is transformed to the RDFS/OWL representation through an XSLT script. The script transforms core components into RDF descriptions that are subclasses of the following OWL types: owl:class, owl:objectProperty and owl:datatypeProperty (see Fig. 3). Unique URIs for RDF descriptions are generated from CCTS names of core components (qualifiers, object class terms, property and representation terms).

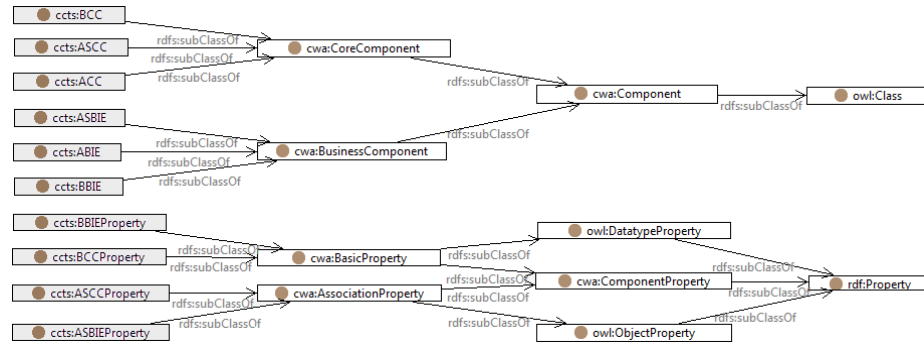


Fig. 3. CCTS artefacts represented as OWL classes. The classes showed on grey background are not in the model designed for validation, but may be used for more precise representation of the CCTS model.

Core component cardinalities are represented in each corresponding OWL class as property restrictions. CCTS cardinalities maxOccurs and minOccurs are represented as OWL properties: owl:maxCardinality, owl:minCardinality and owl:cardinality in conjunction with owl:Restriction as shown in Table 1.

Table 1. Core components cardinalities represented as RDF/OWL restrictions.

maxOccurs	minOccurs	Restriction	owl:Restriction
1	1	exactly 1	owl:cardinality
unbounded	k	Min k	owl:minCardinality
1	0	Max 1	owl:maxCardinality
n	k	Max n, Min k	owl:maxCardinality, owl:minCardinality

In our model, most core data types are transformed into corresponding XML schema data types. However, if a data type contains additional elements (data type attributes in CCTS such as code list identifier in the Code data type), it is represented as a class. Possible values for basic and associated core components are defined in properties as `rdfs:range`.

Definitions and examples for the core components are represented through the introduction of properties `ccts:definition` and `ccts:example`. Additional properties for useful information could be added accordingly. The resulting model is stored as a business ontology for its future use in the validation service.

It is also possible to create a CCTS specific OWL model which contains exactly the same information as the original CCTS model [2]. This type of model can be used for the construction of core components and business documents, including the automatic generation of XML schemas. However, in the case of XHTML+RDFa validation such model is not necessarily needed.

4 Validation of XHTML+RDFa business documents

Social services processes include a substantial amount of processing rules, part of which is visible on the document level. For example, there can be conditionally mandatory parts of the document, with the condition defined by code values. Documents could also include some unconditionally mandatory fields, such as the social security number of the customer. Most of such rules and checks can be implemented on the form level. However, as there could be several different implementations of the same forms, it must be ensured that invalid documents cannot be submitted to the archive. A validation service is a part of the archive that checks incoming documents against a number of rules. In addition to the standard validation of the document structure against its schema, the validation service might check the following:

1. Correctness of code values and code lists used in the document (similar to UBL code list value validation [14])
2. Conditionally mandatory fields and other rule-based checks
3. Adherence to the basic formatting requirements

The validation of business documents is based on the closed world assumption [15] because a document is in fact a snapshot of the information collected on a certain clearly defined issue at a single moment. Although more information may be available elsewhere, it is out of the scope of the document and should not be considered in validation. There have been several proposals for adding integrity constraints into OWL by defining semantics for the constraints [10, 16]. Standard OWL reasoning cannot be used for strict validation because restrictions are used for inferring new information rather than for checking integrity constraint violations.

Our solution for the validation of XHTML+RDFa business documents works as follows:

1. validate the XHTML structure

2. extract the RDF graph from the XHTML markup enriched with RDFa attributes
3. compare the resulting graph with the corresponding RDFS/OWL schema
4. perform additional rule-based checks.

For our purposes, a top-level metamodel is created to separate the classes and properties used for validation (classes in namespace *cwa* in Fig. 3) from standard OWL structures. The metamodel defines new types for classes and properties that are used in validation to check whether the extracted RDF graph is consistent with the corresponding RDFS/OWL schema. The closed world constraint validation service depicted in Fig. 4 can check XHTML+RDFa documents for unknown predicates and classes, unexpected namespaces, ill-formed literals and cardinality constraints.

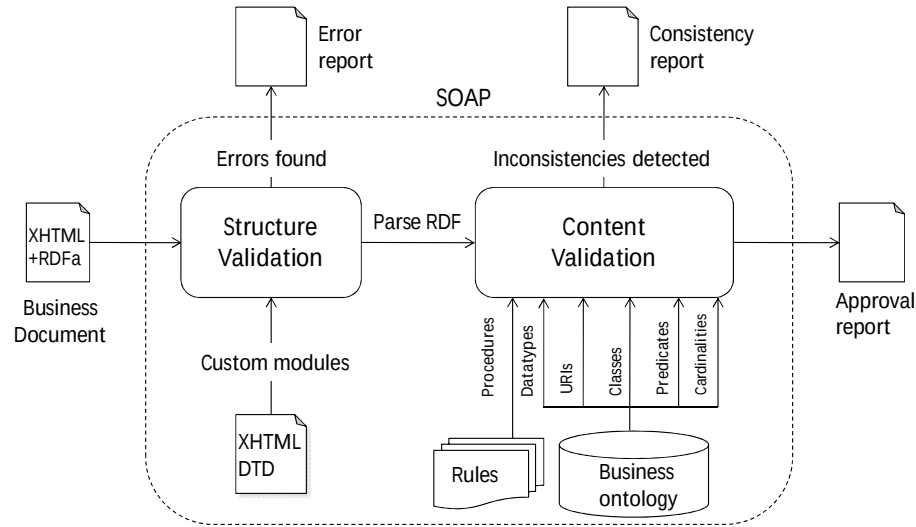


Fig. 4. Closed world constraint validation service.

XHTML is based on a modular framework [9] that can be used to restrict or extend the language. The document can be validated against a custom schema or document type definition (DTD) for XHTML. For instance, the scripting module can be left out, as it is rarely relevant for business documents. In addition, the requirements for archiving might prohibit any references to external resources, as these can get outdated after the document has been archived.

An XHTML+RDFa document is first sent to the validation service wrapped in a SOAP request. The document is then validated against the (potentially customised) XHTML+RDFa DTD. For documents with impaired basic XHTML structure, a SOAP response with relevant error details is generated.

After the validation of the basic document structure, the document is accepted for the content validation, where the embedded RDF graph is extracted

from it. The extracted RDF triples are then inspected against the business ontology which contains the core components, business information entities and business document structures.

The content validation is based on a set of inferencing rules and custom modules implemented using the Jena framework [7]. The triples extracted from the document are merged with the business ontology, and the validation rules are applied to the resulting model through inferencing. The inferencing rules for validation are implemented in the Jena rules language. Most rules are designed to detect incorrect references between components (such as using a wrong BBIE in an ABIE), check the cardinalities and data types of literals, but there is also a number of more general rules, such as checks for unknown namespaces. As an example, the following rule is used for data type checks:

```
[DatatypeRangeError:
  (?documentFragment cwa:isFragmentOf ?s)
  (?s ?p ?o)
  (?p rdf:type cwa:BasicProperty)
  (?p rdfs:range ?range)
  notCastableAs(?o, ?range)
  makeSkolem(?errorNode,?p,?o,?range)
->
  (?errorNode rdf:type cwa:Failure)
  (?errorNode cwa:FailureMessage "Datatype error")
  (?errorNode cwa:onProperty ?p)
  (?errorNode cwa:found ?o)
  (?errorNode cwa:expecting ?range)
]
```

This rule infers several new triples describing a node of type `cwa:Failure` if the content is based on a wrong data type. For example, an ill-formed date in a business document would create the following new triples:

```
_:X1 cwa:expecting xsd:date
_:X1 cwa:found '2011-AUGUST-02'
_:X1 cwa:onProperty cwa:endDate
_:X1 cwa:FailureMessage 'Datatype error'
_:X1 rdf:type cwa:Failure
```

At the end of the validation process, the inferred model is queried for failures. A report is constructed and returned as a SOAP response. The report contains either a list of failures or the approval of the document. A business document is considered valid if its basic XHTML structure is correct and the document is semantically consistent.

5 Discussion

Compared with document exchange based on document type specific XML schemas, our approach has a number of advantages. We list the advantages and illustrate their impact on document exchange and document management.

First, layout information is integrated in the document. There is no need to develop and maintain separate scripts (e.g., XSLT) for producing human-readable versions of the documents. Instead, documents can be viewed in any web browser. At the same time, separate CSS style sheets can be used for fine-tuning the layout, or additional scripts written for converting the documents to other formats, such as PDF/A, which is better suitable for long-time archiving.

Second, all documents are based on a single schema (XHTML+RDFa DTD). This simplifies the management of documents in a single archive used by a number of different information systems. In the Finnish social services system the use of separate schemas for separate document types would mean that the documents in the archive would be based on some 200 schemas, not counting their versions. It is a strong requirement for a single information system to implement support even for a part of them. In the case of a single schema, the system will be able to show the document to the user without resorting to display format generation services, even if it cannot process the semantic markup.

Third, the version management of document structures is easier, because semantics are clearly separated from the document container. A new version of a document does not require a new version of a schema. In a purely schema-based system a new version of a commonly used core component may influence a number of documents and, in turn, force the generation of new versions for their schemas. We note that the management of the core component library itself does not become easier. However, there is no need to keep track of schema namespace changes spawned by the versioning of business information entities.

Finally, semantic markup can be applied only to the information that has clear impact on the automatic processing of information. We noticed that in some social services processes the amount of such information is rather low. This information can be stored as pure XHTML, and semantic markup may be added gradually in the future if the need arises.

A drawback of this approach is that the validation of the document structures becomes more difficult. In the previous section, however, we showed how this drawback can be addressed with closed world validation based on semantic technologies.

There is also a number of new requirements placed on the information systems that need to exchange information. They must produce documents with XHTML+RDFa markup instead of somewhat more straightforward generation of simple XML document instances. Importing the information from a document is also different, as it is not based on parsing the XML structure. Instead, the RDF graph stored in the document must be processed and imported. We note, however, that tool support for such functionality is currently sufficient. Furthermore, the extracted RDF graph can be even transformed to a more traditional XML representation, usable in legacy applications.

6 Conclusion and future work

When there is need for the interoperability of several systems, such as customer information systems, document type specific XML schemas are often used. However, if the number of schemas is big, or schemas are updated often, their management might become troublesome. In addition, scripts (e.g., XSLT) for producing human-readable versions of the documents and other schema-specific infrastructure have to be maintained. In the paper, we present a solution for exchanging CCTS-modelled documents using the XHTML+RDFa document format. In social services applications, where the exchange of human-readable content is sufficient in many cases, our solution can facilitate the incremental addition of semantic markup to the documents. Still, the validation of semantic markup which represents CCTS artefacts is possible to aid the interoperability of different systems.

The feedback received from customer information system developers was more in favour of the traditional schema-based approach. However, the system developers found some aspects of XHTML+RDFa, such as the change management model and the possibility for gradual addition of semantic markup, beneficial. The impact of the proposed solution on other aspects of the social services IT infrastructure has still to be analysed in more detail. We assume that the switch to XHTML+RDFa influences the design of at least the following services: end-user interfaces, code list server, digital signatures, statistics and research services, and message communication protocols. Currently, the use of XHTML+RDFa in Finnish social services is still under consideration.

We implemented a prototype of a validation service for XHTML+RDFa documents with CCTS-based content. The service reports problems in both the basic XHTML structure and CCTS content. We plan to develop the service further to improve its report generator, as currently the service does not indicate the line numbers in which ill-formed RDFa attributes and values are located. In addition, we plan to investigate the possibilities of using XHTML+RDFa documents with CCTS-based content in the implementation of business rules.

Acknowledgement This work was supported by the Tikesos project (National Project for IT in Social Services).

References

1. Adida, B., Birbeck, M., McCarron, S., Pemberton, S.: RDFa in XHTML: Syntax and processing. a collection of attributes and processing rules for extending XHTML to support RDF (2008), <http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014>
2. Biersteker, H., Hodgson, R.: The Netherlands Ministry of Justice Metadata Workbench: Composing XML message schemas from OWL models. Enterprise Data Journal (May 2010), <http://www.enterprisedatajournal.com/article/netherlands-ministry-justice-metadata-workbench-composing-xml-message-schemas-owl-models.htm>

3. Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I.: The semantic web: the roles of XML and RDF. *Internet Computing*, IEEE 4(5), 63–73 (2000)
4. Eriksson, H.: The semantic-document approach to combining documents and ontologies. *International Journal of Human-Computer Studies* 65(7), 624–639 (July 2007)
5. Farrell, J., Lausen, H.: Semantic annotations for WSDL and XML schema (2007), <http://www.w3.org/TR/2007/REC-sawSDL-20070828/>
6. Finnish Government: Archives Act (831/1994 Arkistolaki) (1994)
7. Jena: A semantic web framework for Java, <http://jena.sourceforge.net/>
8. Laudi, A.: The semantic interoperability centre europe: Reuse and the negotiation of meaning. In: Charalabidis, Y. (ed.) *Interoperability in Digital Public Services and Administration: Bridging E-Government and E-Business*, pp. 144–161. IGI Global (2010)
9. McCarron, S., Ishikawa, M.: XHTML 1.1 - module-based XHTML - second edition. W3C Recommendation (November 2010), <http://www.w3.org/TR/xhtml11/>
10. Motik, B., Horrocks, I., Sattler, U.: Bridging the gap between OWL and relational databases. *Web Semantics: Science, Services and Agents on the World Wide Web* 7(2), 74–89 (2009)
11. Mykkänen, J., Hyppönen, K., Kortelainen, P., Lehmuskoski, A., Hotti, V.: National interoperability approach for social services information management in Finland. In: Charalabidis, Y. (ed.) *Interoperability in Digital Public Services and Administration: Bridging E-Government and E-Business*, pp. 254–278. IGI Global (2010)
12. Nešić, S.: Semantic document model to enhance data and knowledge interoperability. In: Devedžić, V., Gašević, D. (eds.) *Web 2.0 & Semantic Web*, vol. 6, pp. 135–160. Springer US (2009)
13. NIEM Technical Architecture Committee (NTAC): National information exchange model naming and design rules. Version 1.3 (October 2008), <http://www.niem.gov/pdf/NIEM-NDR-1-3.pdf>
14. OASIS Universal Business Language (UBL) TC: Universal Business Language v2.0 (December 2006), <http://docs.oasis-open.org/ubl/os-UBL-2.0/>
15. Reiter, R.: On closed world data bases. In: Gaillaire, H., Minker, J. (eds.) *Logic and Data Bases*, pp. 55–76. Plenum Press, New York (1978)
16. Sirin, E., Tao, J.: Towards integrity constraints in OWL. In: *Proceedings of the Workshop on OWL: Experiences and Directions, OWLED 2009* (2009)
17. Thompson, H.S., Beech, D., Maloney, M., Mendelsohn, N.: XML schema part 1: Structures second edition. W3C Recommendation (October 2004), <http://www.w3.org/TR/xmlschema-1/>
18. United Nations. Centre for Trade Facilitation and Electronic Business: Core components technical specification. Version 3.0 (September 2009), <http://unece.org/cefact/codesfortrade/CCTS/CCTS-Version3.pdf>
19. United Nations. Centre for Trade Facilitation and Electronic Business: XML naming and design rules technical specification. Version 3.0 (December 2009), <http://unece.org/cefact/xml/UNCEFACT+XML+NDR+V3p0.pdf>
20. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web* 4, 14–28 (2006)

3. Empowering Citizens and Government with Collaboration on Linked Open Data

Empowering Citizens and Government with Collaboration on Linked Open Data

Kelli de Faria Cordeiro¹, Maria Luiza Machado Campos¹, Marcos R. S. Borges¹

¹ Graduate Program in Informatics (PPGI) - Institute of Mathematics (IM)
Federal University of Rio de Janeiro (UFRJ), Brazil
{kelli, mluiza}@ufrj.br, mborges@nce.ufrj.br

Abstract. Encouraged by many e-Government programs, the volume of Linked Open Data (LOD) is growing every day. With all these data, the citizens and industry are having a great opportunity to be aware of their community and to have an active participation in Governmental decisions. However, transforming data into useful information requires data to be relevant, reliable and interrelated, making the process of publishing data a complex and challenging task. Many automatic approaches are being used to address that, but still with little citizen intervention. In addition to these automatic processes, we propose a collaborative environment that is sensitive and adaptative to citizens' opinions and dynamic needs. It supports several types of direct or indirect collaborations, over linked data elements, including individual contributions, comments mined from Social Networks or issues discussed in forums. Moreover, an exploration interface is designed to provide analytical data to decision-making processes, empowering citizens and improving e-government initiatives.

Keywords: Linked Data, Collaboration, Social Web, Governance, Citizenship

1. Citizenship and Open Governmental Data

The Web of Data, mainly composed of Open Governmental Data (43%) [1], is a powerful environment, which can be used to provide useful information to citizens about their community and decisions involved. Governments are making considerable efforts to open and link their data [2][3] following the standards and principles of Linked Open Data (LOD) [1]. Most of them have initiatives to encourage the participation of citizens in government, as is the case, for example of the Cockpit [4] and Crossroad [5] projects, supported by FP7¹ of the European Commission. The LOD2 Project [6], also supported by FP7, among other issues, prioritizes the development of adaptive tools for authoring and methods for tracking provenance, important requirements for quality and trust on LOD.

The collaboration of citizens within the process of publishing LOD can make the web of data an adaptive environment, more appropriate to their needs, where the content is driven also according to the citizens' interests. The role of citizens in this process depends on the collaboration type, as described in the following section.

¹ FP7 - Seventh Framework Programme. <http://cordis.europa.eu/fp7>

2. Collaboration Types on Governmental Linked Open Data

The citizens' role on complementing the process of linked data publication depends on the type of collaboration, which may vary from a single comment on a data item value to a discussion about the meaning of a concept. They can be classified as following:

- **Direct**: includes the suggestion of a new link [7], an annotation of a data item, a comment or validation on existing links [8], feeding of a new data item (e.g., a picture) [9] and extra links to it [10][11], requests to a government organization to publish some data, and describing geographic places using explicit semantic properties [12];
- **Indirect**: refers to mined information automatically extracted from Web 2.0 mass collaboration applications, such as Social Networks [13][14][15][10];
- **Interaction**: as both previous are visible to citizens, an interactive environment can be provided to support discussion and deliberation [8], for example, about the meaning of concepts to achieve semantic consensus or to expose divergent views of a particular domain, suggestions to develop or improve applications consuming linked data and analyses of aggregated linked data.

In a collaborative environment, citizens actively participate, helping to improve the semantic quality of the open data and links, and to influence the directions to where the Web of Data will evolve. Consequently, government organizations will have immediate feedback, and will become aware of community expectations. More relevant and consistent data might be available, leading to decision-making processes sensible to citizens' demands, considering both geographic and temporal variations.

Clearly, all these volunteer contributions have to be managed [16] and supported by adequate tools. Besides helping interaction and collaboration, control mechanisms, specially designed for Linked Data, are required to deal with authoring tracking [7][9] and provenance capturing [17]. All these features comprise a comprehensive environment for collaboration over linked open data.

3. Collaborative Environment for Governmental Linked Open Data

The Collaborative Environment for Governmental Linked Open Data (Figure 1) is composed of (i) **Publication Elements**, which transform raw data into Linked Data using wrappers or other transformation mechanisms; (ii) **Consumption elements**, which explore the exposed linked data using browsers, search engines, mashups and applications to support decision making processes; (iii) **Data**, representing instances and schemas, and **Links**, connecting them, all in RDF format; and (iv) **Collaboration elements**, which support different types of collaboration interactions. Besides, a **Semantic Layer** supports all these elements, taking care of the open data and links meaning. Direct users of this environment are **Citizens and Industry** that should have an active voice on the available resources, helping **Governmental Organizations** to improve e-government services.

The Collaboration elements support the interactions between community and government, offering several entry points. Each one has an interface for reliable sources, specially designed to receive information with high level of trust, and an interface for volunteers with none or little control. In this last case, the main goal is to stimulate participation without formal restrictions. On both types of interaction, control mechanisms should automatically collect georeferenced location, timestamp and MAC address of the input device. These data should feed engines for accreditation, authoring control and provenance capturing, supporting the identification of collaborators and the classification of data quality and trust. These data may also be used to dynamically rank the reputation of collaborators over time and to filter or identify information quality and provenance in mashup applications.

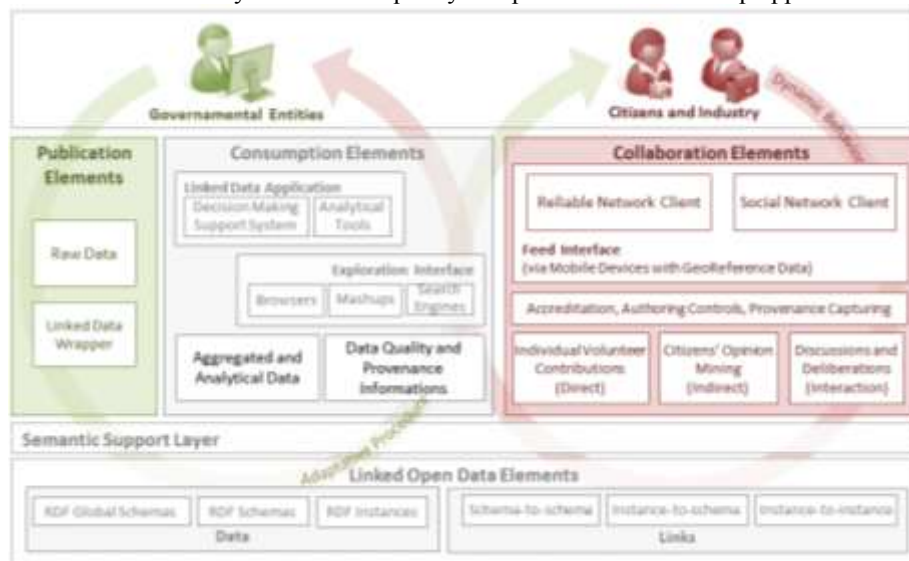


Figure 1 – Collaborative Environment for Governmental Linked Open Data

To handle and support different types of collaboration, there are components with different features and controls to get information (i) directly from individual contributions, (ii) indirectly through mined data from social networks, and (iii) from tools that support interactions over linked data elements. For proper referencing, it is necessary to identify each Linked Data Element indicated in all interactions. Mechanisms are also required to visualize data connections, both at instance and schema level.

Each type of link has different collaboration requirements. For example, identification of generalization and specialization relations between entities requires interactions among users, and this does not happen in individual contributions. Finally, all data on this environment might be aggregated and derived as analytical information to support decision-making processes, empowering citizens and industry, but also improving e-government initiatives. Both citizens and government can feed the data cloud with new collaborations and linking new data, making the LOD cloud an adaptative and evaluative environment, varying according to citizens and industry dynamics.

4. Conclusions

This work proposes an environment to empower citizens, industry and government with collaborative linking of Open Data, by orchestrating several components in order to provide the means to capture community opinions and to analyze them. For this, the Collaboration Types and Linked Data Environment Elements have been identified to support the development of several components, especially the trust and exploration mechanisms that allow consumption of both data and links contributions. The analysis of these contributions supports decisions about the environment, for example, on opening new data or on official confirmation of new links, creating a solid foundation for community governance, empowering both citizens and government.

References

1. Heath, T., Bizer, C.: Linked Data - Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool (2011)
2. UK Government Data, <http://data.gov.uk>
3. Spain GeoLinkedData, <http://geo.linkeddata.es>
4. COCKPIT - Citizens Collaboration and Co-creation in Public Service Delivery, <http://www.cockpit-project.eu>
5. CROSSROAD - A Participative Roadmap for ICT Research in Electronic Governance and Policy Modelling, <http://crossroad.epu.ntua.gr>
6. LOD2 - Linked Open Data 2 Project, <http://lod2.eu>
7. Halb, W., Raimond, Y., Hausenblas, M.: Building Linked Data for Both Humans and Machines. LDOW. Beijing, China (2008)
8. Luczak-Rösch, M., Heese, R.: Linked Data Authoring for NonExperts. In: Linked Data on the Web Workshop, 18th International World Wide Web Conference. Madrid, Spain (2009)
9. Geobuddies - Anotación semántica colaborativa con dispositivos móviles en el Camino de Santiago, <http://www.geobuddies.net/>
10. Tramp, S., Heino, N., Auer, S., Frischmuth, P.: Making the Semantic Data Web Easily Writeable with RDFauthor. LNCS, vol. 6089, pp. 436--440. Springer, Heidelberg (2010)
11. Cordeiro, K.F., Marino, T., Campos, M.L.M., Borges, M.R.S.: Use of Linked Data in the Design of Information Infrastructure for Collaborative Emergency Management System, To be published in 15th CSCWD. Lausanne, Suisse (2011)
12. Braun, M.: Context-aware collaborative creation of semantic points of interest as Linked Data. Master's thesis, University of Koblenz-Landau, Germany (2009)
13. Shvartzshnaider, Y.: Global Semantic Graph as an Alternative Information and Collaboration Infrastructure. ESWC. Heraklion, Crete, Greece (2010)
14. Choudhury, S., Breslin, J., Passant, A. Enrichment and ranking of the youtube tag space and integration with the linked data cloud. ISWC. LNCS. Springer (2009)
15. Duchateau, F., Hardman, L.: Integrating and Ranking Interests From User. Workshop Profiles, ESWC. Heraklion, Crete, Greece (2010)
16. Chklovski, T., Gil, Y.: Towards Managing Knowledge Collection from Volunteer Contributors. Symposium on Knowledge Collection from Volunteer Contributors. (2005)
17. Flouris, G., Fundulaki, I., Pediaditis, P., Theoharis, Y.: Coloring RDF Triples to Capture Provenance. ISWC. USA (2009)

4. Publishing Linked Data from Brazilian Politicians on the Web

Publishing Linked Data from Brazilian Politicians on the Web

Lucas de Ramos Araújo¹, Pablo N. Mendes², Jairo Francisco de Souza¹

¹Knowledge Engineering Research Center (NEnC), Department of Computer Science, Federal University of Juiz de Fora (UFJF), Brazil

²Web-based Systems Group (WBSG), Freie Universität Berlin, Germany
lucas.araujo@ice.ufjf.br, pablo.mendes@fu-berlin.de, jairo.souza@ufjf.edu.br

Abstract. Since its inception, the Web has undergone continuous evolution in order to improve itself as a means of global communication and information sharing. Open Government Data are increasingly being published on the Web, contributing to the transparency and the reusability of public data. At the same time, the use of Linked Data has been increasing in recent years, enabling the development of better and smarter applications. In this context, this paper proposes a citizen-driven initiative for the publication and promotion of Open Government Data from Brazil as Linked Open Data. We create a data set of Brazilian politicians with information collected from different sources and discuss information exploration scenarios that are enabled by this new set of linked data. We expect this dataset to serve as one of many bi-directional links between government and citizen opinion.

Keywords: Linked data, e-government, semantic web, open government data.

1. Introduction

Information and Communication Technologies have promoted a revolution in the information sharing, building a new relationship between government and citizens. This new relationship originated the so-called Electronic Government, which allows a more accessible, efficient, democratic and transparent government. The concept of Open Government Data (OGD) extends this relationship by promoting the availability of government information in open and accessible formats to enable reuse and interconnection between information from different sources, thus generating new knowledge [1].

Nowadays, much government data are available on the Web, but this information is usually offered without the use of standards, in proprietary formats and structured only for user interface presentation, making it difficult for third-party systems to automatically reuse them. For better exploiting the potential represented by the government information, the data must be available in standard, open and accessible formats [2].

Linked Open Data (LOD) is used to describe a set of practices to publish, share and connect structured data on the Web in order to increase the value and usefulness of these data [3]. There is a global movement of governments, organizations and people publishing Open Government Data. At the same time, the use of Linked Data is strongly supported by the W3C¹ (World Wide Web Consortium) as well as the academic community and industry².

The importance of publishing Open Government Data and using Linked Data on the Web nowadays has become evident. Open Government Data projects are emerging in different countries around the world, such as the U.S., U.K., Australia, New Zealand, Norway, Holland, Sweden, Spain, Austria, Greece, Canada and Denmark. There are also a growing number of local initiatives from states and cities [4]. Some governments have created catalogs or portals to facilitate the location and use of data for the public [5]. In addition, individuals and organizations are publishing government data on their own in various formats [6].

Brazil has a substantial amount of government data available on the Web, but there are only a few government initiatives that give full access to these data in structured and open language. Therefore, some initiatives are emerging in order to extract the data, make them open and give them new value through different applications. According to the United Nations E-Government Survey 2010 [7], which provides a global assessment in the field of Electronic Government, Brazil ranks 61st, losing 16 positions since 2008. Several factors are responsible for the decline, such as the lack of online services and poor telecommunications infrastructure. The report also highlights Brazilian initiatives on Open Government Data that must be followed.

This paper aims to contribute in both Open Government Data and Linked Open Data fronts by publishing Linked Data from Brazilian politicians on the Web using information collected from different sources. Our work aims to provide useful, open, standardized, reusable, and linked data for both human and machine consumption. As a result, new kinds of information exploration become possible. Due to the flexibility offered by the RDF data model, users can easily take a politician-centered view, or a geographically centered (geocentric) view on the information. Due to the use of standard vocabularies, citizens can issue queries comparing indicators from Brazil with those of other governments. And due to the integration with external sources, users can ask a richer variety of questions including, for example, demographics, social indicators or other encyclopedic knowledge associated with their regions.

According to the Latin American Public Opinion (LAPOP) [8], there is a great lack of interest in politics among the Brazilian citizens: over 90% of the interviewed people have average, little or no interest in it. Also, according to a survey conducted by the Brazilian Institute of Geography and Statistics (IBGE) [9], 28% of those interviewed did not know the name of the governor of his state, while 17% did not know the name of the President. The contributions presented in this paper are a first step towards engaging citizens in actively participating in the decisions being taken by their elected government.

¹ <http://www.w3.org/standards/semanticweb/data>

² <http://linkeddata.org>

The remainder of this paper is structured as follows: Section 2 provides an overview on Open Government Data and Linked Data. Section 3 presents information regarding the implemented project. Section 4 presents the overall architecture and data generation. Section 5 presents a discussion on the impact of our work on citizen engagement with the government, while Section 6 discusses conclusions and future work.

2. Background

According to the E-Government Interest Group – e-Gov IG [10], creating an Electronic Government requires openness, transparency, collaboration and knowledge. A transparent government is more than open interaction and participation; government data must be shared, discovered, accessible and manipulated by those who wish use government information.

The aim of Open Government Data is to allow users to easily find, access, understand and use public data, bringing many benefits such as reuse, inclusion, transparency, accountability, improved search, integration, participation, collaboration, economic growth, innovation and efficiency [2]. The Open Government Working Group [11] developed the 8 Principles of Open Government Data. They should be: Complete, fully available without limitations; Primary, collected at the source with the finest possible level of granularity, without aggregation or modification; Timely, published as quickly as necessary to preserve its value; Accessible, available to the largest possible number of users and purposes; Machine Processable, reasonably structured to allow automated processing; Non-discriminatory, available to anyone without registration; Non-proprietary, available in a format over which no entity has exclusive control; and License-free, not subject to any copyrights, patents, intellectual property or trade secrets regulation.

The main technologies and formats used to publish Open Government Data are: CSV (Comma Separated-Values) files, which store tabular data; Atom and RSS (Really Simple Syndication) information, which aggregate content based on XML (Extensible Markup Language); REST interfaces, which link a resource to an URI using HTTP (Hypertext Transfer Protocol), allowing a site to be enriched with applications that extend the value of an available resource; and Semantic Web and Linked Data technologies, which offer a common framework where data can be shared [10].

Linked Data refers to data published on the Web in a way that they facilitate programmatic processing and linking to/from external data sets [3]. The basic idea was developed by Berners-Lee [12]. He defined the four principles that characterize Linked Data and that should be applied to make the Web grow: use URIs for naming “things”; use HTTP URIs so that people can look up these names; provide useful information when someone looks up a URI; and include links to other URIs so that people can find more related “things”.

According to Berners-Lee [6], the overall benefits of Linked Data are: they are accessible by an unlimited variety of applications because they are expressed in open formats; they can be combined through mash-ups with any other set of Linked Data, it is easier to integrate these data sources as long as they both use Linked Data standards; it is easy to add more Linked Data to the existing ones, even when the terms and definitions change over time.

RDF links are the foundation of Linked Data. They allow client applications to navigate between data sources and find additional data. Thus, data sources must set RDF links to relate the entities in other data sources [3]. RDF links can be set manually, but it is a common practice to use automated or semi-automated approaches to generate RDF links [13]. For client applications to process data more easily, it is considered a good practice to reuse terms of well known and widely used vocabularies whenever possible. New terms should only be defined if the required terms cannot be found in existing vocabularies. In this case, a file must be published with the new terms created using RDFS or OWL [6]. To ensure that consumers can evaluate the quality of the published data and determine if they want to trust them, the data must be accompanied by various types of metadata, such as a URI that identifies the author, the creation date and the creation method [14].

The most visible example of the adoption and application of the Linked Data principles is the *Linking Open Data*³ project, an open and collaborative effort supported by the W3C SWEO (Semantic Web Education and Outreach) Group. Its goal is to identify existing data sets available under open licenses, convert them to RDF according to the principles of Linked Data, publish them on the Web and link them with each other, generating a LOD Cloud (Linking Open Data Cloud) [14].

In Brazil there are few government initiatives that give full access to the data in a structured and open language. Examples of these initiatives are the projects *Governo Aberto SP*⁴ and *LeXML*⁵. While the government does not release more data in open format, initiatives are emerging in order to extract data from government sites and portals, rearrange them, make them open and/or provide new value to them through different applications, such as the projects *Congresso Aberto*⁶, *Parlamento Aberto*⁷, *Legisdados*⁸ and *SACSP*⁹. Due to the few Open Government Data projects in Brazil, there are still not many projects that mix information from different sources to create new services and offer new insights about the government. Our project aims at contributing to fill this gap.

³ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁴ <http://www.governoaberto.sp.gov.br/>

⁵ <http://www.lexml.gov.br/>

⁶ <http://www.congressoaberto.com.br/>

⁷ <http://trac.meuparlamento.org/>

⁸ <http://www.legisdados.org/>

⁹ <http://sacsp.mamulti.com/>

3. Brazilian Politicians data set

The aim of our work is to provide a new data set with information from Brazilian politicians gathered from different sources using the Linked Data and Open Government Data practices. The name “Ligado nos Políticos” – literally, “linked to the politicians”; figuratively, “paying attention to the politicians” – and the domain “<http://ligadonospoliticos.com.br>” were used to represent the data set and publish the data on the Web. Figure 1 presents the general architecture of our work.

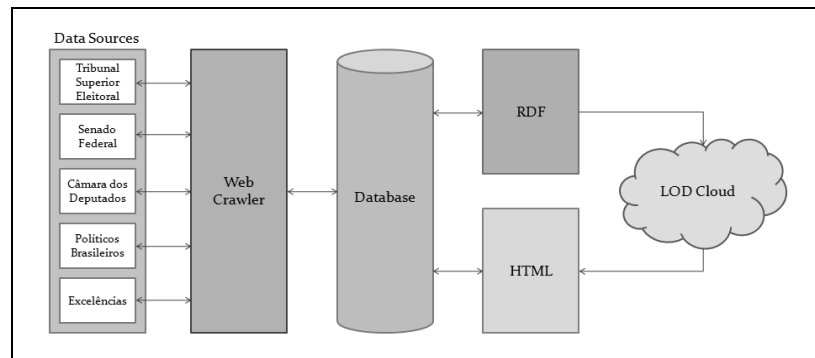


Fig. 1. General architecture of our work

With the exception of some data in CSV format, most of the data sources offer information only in HTML format. Therefore it was necessary to create and use Web Crawlers to extract data in a methodical and automated way. The technique for extracting structured data from HTML pages is called screen scraping. There are several tools to assist in this practice, in different languages and platforms. In this project the language PHP (Hypertext Preprocessor) and XPath expressions were used to create custom scripts to extract information from each data source.

The data sources used were the following websites: *Tribunal Superior Eleitoral*¹⁰ (Supreme Electoral Court), *Senado Federal*¹¹ (Federal Senate), *Câmara dos Deputados*¹² (House of Representatives), *Políticos Brasileiros*¹³ (Brazilian Politicians) and *Excelências*¹⁴ (Excellencies). Personal data, election data, disclosure of assets, parliamentary data, leaderships, missions, mandates, clearances, speeches, commissions, proposals, legal occurrences and votes on bills were collected. At the same time, data regarding the extraction and the provenance of the data were also collected, such as the website, the URL, the date of extraction and the hour of extraction. An overview of the data collected from each data source is provided in Table 1.

¹⁰ <http://www.tse.gov.br/>

¹¹ <http://www.senado.gov.br/>

¹² <http://www.camara.gov.br/>

¹³ <http://politicosbrasil.org.br/>

¹⁴ <http://www.excelencias.org.br/>

Data Source	Data Collected
Supreme Electoral Court	Personal Data (full name, date of birth, nationality, place of birth, education level, gender, marital status, homepage and occupation), Declaration of Assets (description, type and value) and Election Data (party, alliance, office, campaign spending limits, campaign protocol and campaign number).
Federal Senate	Personal Data (full name, date of birth, place of birth), Political Data (governmental name, office, party, state), Governmental Address (wing, cabinet, street, neighborhood, city, state, zip code, email, phone, fax, homepage), Leaderships (description, start date, end date), Missions (description, start date, end date, document), Propositions (title, description, autor), Mandates, Absence, Comissions (description, start date, end date, participation) and Speeches (type, date, house, party, state, summary).
House of Representatives	Personal Data (full name, date of birth, profession), Political Data (office, party, state), Governmental Address (annex, cabinet, street, neighborhood, city, state, zip code, CNPJ, email, phone, fax, homepage) and Speeches (date, session, period, hour, summary).
Brazilian Politicians	Personal Data (full name, gender, date of birth, marital status, nationality, place of birth, education level, occupation) and Political Data (office, party, state).
Excellencies	Legal Occurrences (title, acronym, type, year, number, description).

Table 1. Overview of the data collected from each data source

There was not a common identifier between the different data sources, due to the fact that a key is not published alongside the politics data to ensure that they are unique, like the ISBN (International Standard Book Number) for books. The task of how to best detect that two entity descriptions refer to the same real world entity is an open research problem known as duplicate record detection [15]. In our case, we benefited from different attributes from politicians, such as name, party and date of birth to identify the data that addressed the same entity and ensure consistency and avoid duplication, using the different data at the WHERE clauses before inserting a new entry at the database. Moreover, some semi-automated data cleansing was performed with UPDATE queries. For instance, reconciling different spellings of party names (PC do B versus PCdoB) or normalizing state names to acronyms (Rio de Janeiro versus RJ).

After having all the structured data, the information was represented using the principles and practices of Linked Data and using the RDF model, Content Negotiation and dereferenceable URIs to ensure that the non-informational resources were correctly generated through good URIs for both HTML and RDF clients. Simple and small HTTP URIs were chosen to uniquely identify the non-informational resources.

Subsequently, in order to describe our dataset, we selected terms from the following vocabularies: rdf, rdfs, owl, foaf, bio, person, vcard, dbpprop, po, pol, being, timeline, dc, spin rdf, skos, earl, geospecies, rdf money and biblio. The *Talis*

*Schemacache*¹⁵ service was used to find the terms, and other data sets were also investigated, such as *GovTrack.us*¹⁶. Not all the terms needed were found in other vocabularies. In such cases, new terms were defined under the namespace “<http://ligadonospoliticos.com.br/politicobr/>” represented by POLBR.

With the vocabulary defined, the information was represented in the RDF model. The information is taken from the database and inserted dynamically into the model according to the requested resource. The subject is represented by the politician URI, the predicates by the URIs from the preexisting or created terms, and the objects are described by literal values taken from the database or by URIs that represent the RDF links to other resources. Besides politicians, the data set contains information about locations, occupation, as well as senate and congress bills voted by politicians.

RDF links to data sets such as DBPedia¹⁷, GeoNames¹⁸, Freebase¹⁹, World Factbook²⁰, UMBEL (Upper Mapping and Binding Exchange Layer)²¹ and YAGO²² were generated for the politicians, groups, cities, states, country, occupations and parties. For this, properties like *owl:sameAs*, *foaf:homepage*, *foaf:page*, *foaf:primaryTopic*, *rdfs:seeAlso*, *rdf:type* and *skos:subject* were used. Since there were no common identifiers between the data sets, manual and semi-automatic approaches were used to generate the links, the latter through practices similar to those used in the Web Crawler, traversing the pages of other data sources, comparing the resources properties and extracting the links. Since the quality of the links generated cannot always be guaranteed, a better approach is being implemented using the Silk Link Discovery Framework²³, which provides a better support in the generation of links between different data sources, specifying which types of RDF links should be discovered as well as which conditions data items must fulfill in order to be interlinked.

In order to facilitate the discovery of data by machines and humans, the data set was added in the *ESW Wiki*²⁴, registered in *CKAN*²⁵ and a sitemap was created to indicate where the RDF is located. To test whether the data can be handled properly and if they can follow links RDF, *The Tabulator Firefox Extension*²⁶ was used (Figure 2). The *W3C RDF Validation Service*²⁷ was also used to ensure that valid RDF/XML documents are provided.

¹⁵ <http://schemacache.test.talis.com/>

¹⁶ <http://www.govtrack.us/>

¹⁷ <http://dbpedia.org/>

¹⁸ <http://www.geonames.org/>

¹⁹ <http://www.freebase.com/>

²⁰ <http://www4.wiwiw.fu-berlin.de/factbook/>

²¹ <http://www.umbel.org/>

²² <http://mpii.de/yago>

²³ <http://www4.wiwiw.fu-berlin.de/bizer/silk/>

²⁴ <http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets>

²⁵ <http://ckan.net/package/brazilian-politicians>

²⁶ <http://dig.csail.mit.edu/2007/tab/>

²⁷ <http://www.w3.org/RDF/Validator/>

DILMA VANA ROUSSEFF	nationality	Brasileira nata
	office	Presidente
	governmentalname	DILMA ROUSSEFF
	maritalstatus	Divorciado
	situation	Candidato
	state of birth	http://www.geonames.org/3457153
	complexion	Branco
	occupation	http://rdf.freebase.com/rdf/en.economist
	created	2010-12-02
	creator	http://ligadonospoliticos.com.br/content/foaf.rdf
	publisher	http://ligadonospoliticos.com.br/content/foaf.rdf
	rights	http://ligadonospoliticos.com.br
	education Level	Superior completo
	language	pt-br
	place of birth	http://dbpedia.org/resource/Belo_Horizonte
		http://www.geonames.org/3470127
	father	Pedro Rousseff (P��tar Russ��v)
	mother	Dilma Jane Silva
	party	PT
	type	http://dbpedia.org/class/yago/BrazilianPoliticians
		http://dbpedia.org/class/yago/LivingPeople
		http://dbpedia.org/ontology/Person
		http://dbpedia.org/ontology/Politician
		Thing
	label	Descri��o RDF de DILMA VANA ROUSSEFF
	see Also	http://twitter.com/MinistraDilma

Fig. 2. RDF representation with The Tabulator Extension

Next, a representation for HTML viewing and searching data was created. The home page provides an overall description, the number of registered politicians, links to other pages on the site, and especially a search engine for users to find the desired politician according to different criteria such as name, location, status, state, party and gender. One or more criteria can be used to perform the search. By selecting a politician, the proper representation is presented according to the HTTP header sent by the client (HTML or RDF). The politician data are dynamically extracted from the database and displayed on this page. To improve the navigability, certain data can be selected (with a click or issuing another HTTP request) to search politicians with the same characteristics.

In order to improve the users experience and allow for analytical insight from the collected data, different visualizations are generated dynamically from queries in the database using *FusionCharts Free*²⁸. We allow people to choose different parameters for personalized queries. The user can, for example, visualize a chart of the number of speeches from male senators from PSDB (party).

A link to RDF/XML format is also presented at the beginning and at the end of the page. To take advantage of the Web of data and the RDF links, data from other sources were also extracted and presented via SPARQL queries using *ARC*²⁹, a flexible RDF system for semantic web using PHP. An example of the HTML representation is provided on Figure 3.

²⁸ <http://www.fusioncharts.com/free/>

²⁹ <http://arc.semsol.org/>



Fig. 3. HTML representation for the Brazilian President “Dilma Rouseff”.

Analyzing our work according to the Open Government Data quality requirements, we can say that the published data generally meet the basic principles that determine this practice: they are complete, as all data collected are available without restrictions of privacy, security or privileges access; they are primary, since they have the finest possible level of granularity without modifications; they are timely, as they are recent data from the Brazilian politicians; they are accessible, available to any Web user; they are machine processable, structured using the RDF model; they are not-discriminatory, because there is no need to register to access the data; they are not-proprietary, as no entity holds the exclusive control of the data, and they are license-free, as they are public data collected from different government sites. Regarding Linked Data, our work also meets its basic principles, using HTTP URIs for naming resources, allowing people to access them, providing useful information using the RDF standard and including links to other URIs so that users can find more resources. With the data set from Brazilian politicians online, it is possible to query the data, extract data published via RDF, link these data with data from other data sets and develop different web applications.

4. Linking Citizens and Politicians

This work is a contribution towards narrowing the gap between citizens and politicians through open Web technologies. This has been done in four different ways: emphasizing the impact of each citizen’s vote, facilitating citizen oversight, empowering citizen opinion and seeding an open government mash-up ecosystem. This section describes these ideas in more detail.

First, we emphasize the interpretation that politicians represent a certain group of citizens, namely the voters that elected a congressman or senator. The collected election data establishes a link from citizens, through their location of voting, to the politicians elected by those locations. A Location Page – e.g. the page for Rio de Janeiro³⁰ – shows a summary of topics (tag cloud) supported by politicians elected by

³⁰ http://www.ligadonospoliticos.com.br/?pag=estado&id_estado=RJ

that location (Figure 4). We intend to use this view to catalyze on the voters a sense of responsibility for the consequences of the actions of their elected representatives.

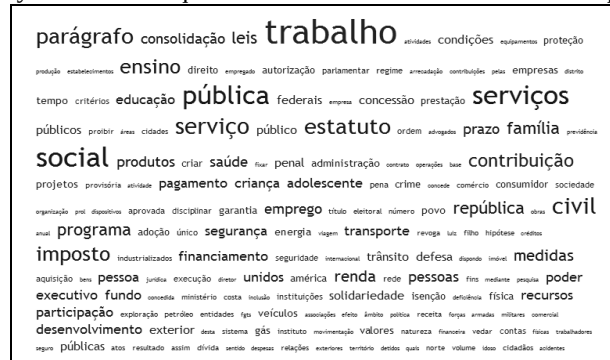


Fig. 4. Tag cloud from bills in the state of Rio de Janeiro showing as prominent terms ‘trabalho’ (work, jobs), ‘social’ (social), ‘serviços’ (services), ‘pública’ (public), etc.

Second, we allow citizens to ask questions about their representatives directly to the data. With the information available as linked data, we have the advantage to query them more flexibly and from different perspectives. Example queries include: the most controversial bills, bills more distant from public opinion, bills in education with high rejection in areas with high illiteracy, bills created by representatives of the poorest state in Brazil, among others. Such direct questions allow citizens to quickly assess the politicians’ impact on key issues, possibly prompting citizen engagement in expressing their satisfaction or disagreement with the work conducted by their representatives.

Third, we allow politicians and citizens alike to follow online discussions that may be relevant with regard to a politician, legislation or location. A preliminary implementation of semantic annotation of citizen opinions allows us to link general public comments to the politician, legislation or location targeted by each comment. For instance, on each politician page we display Twitter posts that mention that politician’s name. Twitter feeds are also collected and displayed using the more common words in bills and speeches. Future work will extend this preliminary work in light of the Linked Open Social Signals architecture [16], enabling more flexibility and expressiveness in selecting a feed of citizen observations.

Finally, a wide variety of applications, mashups and visualizations have been created, especially in the U.S. and the U.K, such as the projects *Where Does My Money Go?*³¹, *FixMyStreet*³², *TheyWorkForYou*³³, among others. These applications offer useful information to citizens, showing the potential of reusing Open Government Data to create new knowledge. As a result of the integration and publication of the Brazilian politicians’ data set as LOD, our project lowers the entry barrier for third-party ‘citizen developers’, researchers or government agencies to

³¹ <http://wheredoesmymoneygo.org/>

³² <http://www.fixmystreet.com/>

³³ <http://www.theyworkforyou.com/>

repurpose government data from Brazil in interesting and useful ways. Moreover, other sources could provide additional information and aggregate data through other perspectives not predicted in this project.

5. Conclusions

The publication of Open Government Data and Linked Data has been increasing in recent years. Still, much remains to be done to evolve the Web of documents to a Web of data accessible to all. Other countries should follow the examples of the U.S., U.K. and E.U. to ensure that more open data are published from government initiatives. Following their successful examples, catalogs or portals should be prepared to serve as a single point of access to data enabling the creation of new applications on top of that data. Linked Data technologies are a natural candidate for these requirements.

Brazil already publishes government data on the Web, but higher conformance with best practices established by Open Government Data and Linked Data initiatives is needed. That way, more applications, mashups and visualizations can be created to provide useful information to citizens. Moreover, the transparency of the government would be increased, improving the country's ratings with external agencies, possibly generating positive socio-economic impact for Brazil.

At the same time, more citizens and NGOs must publish government data on their own in order to establish a data-driven conversation with their governments. Our project is one step in bridging a gap between the government and public opinion. We collected public data that were isolated in silos, linked related instances from those sources, reused pre-existing vocabularies to describe the data when possible, published them in RDF as linked data, offered interfaces for visualization in different perspectives and automatically connected public opinion through social media monitoring.

As future work, we intend to keep the data up to date, as well as include information from candidates, senators and representatives from past years. More RDF links related to other data sources should also be created through more detailed researches on other data sets. We also intend to provide a more flexible interface for arbitrary querying and visualization, for example through *Cuebee*³⁴, and extract entities from text through entity extraction tools, for example *DBpedia Spotlight*³⁵.

Just as the Electronic Government revolutionized the relationship between citizens and government, Open Government Data has the potential to enhance this relationship. And just as the Web has caused a revolution in publishing and consumption of documents, Linked Data has the potential to improve the way data is accessed and used. If the remaining challenges are addressed properly, these practices will enable an even greater evolution in the way the Web is used today.

³⁴ <http://cuebee.sourceforge.net/>

³⁵ <http://wiki.dbpedia.org/spotlight>

References

1. W3C Brazilian Office: O Governo de inovação na Copa 2014: uso de redes sociais e dados abertos. II Seminar on Electronic Government Innovation. Porto Alegre, Rio Grande do Sul. September, 2010, http://www.procergs.rs.gov.br/uploads/1285856001W3C_Seminario_Inovacao_eGov_POA_17092010.pdf. December, 2010.
2. Agune, R. M. Filho, A. S. G. Bolliger, S. P.: Governo Aberto SP: Disponibilização de Bases de Dados e Informações em Formato Aberto. III Consad, Public Managment Congress, Panel 13/050, 2009, http://www.repositorio.seap.pr.gov.br/.../governo_aberto_sp_disponibilizacao_de_bases_de_dados_e_informacoes_em_formato_aberto.pdf. December, 2010.
3. Bizer, C. Heath, T. Berners-Lee, T.: Linked Data - The Story So Far. Preprint to the Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), 2009, <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. December, 2010.
4. Sheridan, J. Tennison, J.: Linking UK Government Data. LDOW. Raleigh, North Carolina. April, 2010, http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf. December, 2010.
5. Bennett, D. Harvey, A.: Publishing Open Government Data. W3C Working Draft. September, 2009, <http://www.w3.org/TR/gov-data>. December, 2010.
6. Berners-Lee, T.: Putting Government Data Online. Design Issues. Junho, 2009, <http://www.w3.org/DesignIssues/GovData.html>. December, 2010.
7. United Nations: United Nations E-Government Survey 2010. New York: UN Publishing Section, 2010, <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan038851.pdf>. December, 2010.
8. Opinião Pública.: Vol. 15, nº 2, November, 2009, Encarte Tendências, Campinas, Brazil. p.511-542. Latin American Public Opinion (LAPOP). 2007. <http://www.scielo.br/pdf/op/v15n2/09.pdf>. March, 2011.
9. Instituto Brasileiro de Geografia e Estatística (IBGE): Pesquisa Mensal de Emprego. Representação de Interesses e Intermediação Política. Abril, 1996. <http://www.ibge.gov.br/home/estatistica/populacao/condicaoodevida/indicadoresminimos/suppme/analiseresultados1.shtm>. March, 2011.
10. E-Government Interest Group (E-Gov IG): Melhorando o acesso ao governo com o melhor uso da web. Comitê Gestor da Internet no Brasil. 1a edição. São Paulo, 2009, <http://www.w3c.br/divulgacao/pdf/gov-web.pdf>. December, 2010.
11. Opengovdata.Org: Open Government Data Principles. California. December, 2007, http://resource.org/8_principles.html. December, 2010.
12. Berners-Lee, T.: Linked Data. Design Issues. July, 2006, <http://www.w3.org/DesignIssues/LinkedData.html>. December, 2010.
13. Bizer, C. Cyganiak, R. Heath, T. How to Publish Linked Data on the Web". July, 2007, <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial>. December, 2010.
14. Bizer, C. et al.: Linked Data on the Web. LDOW 2008. Workshop at the 17th International World Wide Web Conference. Beijing, China. April, 2008, <http://events.linkedata.org/ldow2008/papers/00-bizer-heath-ldow2008-intro.pdf>. December, 2010.
15. Elmagarmid, P. et al.: Duplicate Record Detection: A Survey". January, 2007, <http://www.cs.purdue.edu/homes/ake/pub/survey2.pdf>. January, 2011.
16. Mendes, P. et al.: Linked Open Social Signals. Web Intelligence. Toronto, Canada. September, 2010. <http://knoesis.wright.edu/library/download/paper-wi10-MPKS.pdf>. March, 2011.

5. Policy Intelligence in the Era of Social Computing: Towards a Cross-Policy Decision Support System

Policy Intelligence in the Era of Social Computing: Towards a Cross-Policy Decision Support System

Riccardo Boero¹, Enrico Ferro², Michele Osella²,
Yannis Charalabidis³, Euripidis Loukis³

¹Regione Piemonte, Corso Regina Margherita, 174 – 10152 Torino, Italy
ricboero@gmail.com

²Istituto Superiore Mario Boella, Via Boggio, 61 – 10138 Torino, Italy
{enrico.ferro, michele.osella}@ismb.it

³University of the Aegean, Gorgyras Str. – 83200 Karlovassi, Greece
{yannisx, eloukis}@aegean.gr

Abstract. The paper presents a policy analysis framework developed through a process of interdisciplinary integration as well as through a process of end users needs elicitation. The proposed framework constitutes the theoretical foundation for the decision support component of a technological platform bringing together Social Media and System Dynamics simulation developed within the PADGETS project. The main novelties introduced have to do with the possibility to provide decision makers with a set of synthetic, fresh and relevant data in a cost effective and easily understandable way.

Keywords: ICT Governance, Policy Intelligence, Policy Modeling, eParticipation, Social Media, Decision Support Systems, Public Sector Innovation, eGovernment, PADGETS.

1 Introduction

In the second decade of the millennium European Governments are confronted with three important long term trends.

1. The combined effect of an increase in the rate of change and in the level of interdependence and interconnectedness among regions, activities and groups is leading to a fast-evolving and unpredictable world characterized by significant levels of complexity and uncertainty. The concept of “liquid modernity” [1] proposed by Zygmunt Bauman represents a useful attempt to frame part of such phenomenon. According to the Polish sociologist, in fact, social forms and institutions no longer have enough time to solidify and cannot serve as frames of reference for human actions and long-term life plans to the extent they served in the past, so individuals have to find other complementary ways to organize their lives.
2. A push towards a more participatory and inclusive style of policy making poses significant challenges in terms of striking the right balance between openness and

control, defining new and appropriate styles of management and, finally, integrating participatory activities into existing decision making processes.

3. A forecast, for the years to come, of low economic growth and financial instability is leading to tighter budget constraints and less room for mistakes in the allocation of tax payers' money for Government's action.

The concurrence of such socioeconomic, institutional and financial trends calls for a reconceptualization of Government's roles and *modi operandi*.

The creation of the PADGETS (its full title being "Policy Gadgets Mashing Underlying Group Knowledge in Web 2.0 Media") project [2] may be placed in such scenario. The project has been financed in the context of the "ICT for Governance and Policy Modelling" call of the seventh European Framework Program of research (FP7). The main underlying idea of such research endeavor is to bring together social computing with System Dynamics simulation in order to help Governments to render policy making processes more participative and, at the same time, to provide advanced and more effective types of support to public sector decision making processes. In particular, the platform developed within the project will allow Public Administrations to set up a cost effective participatory process by moving the political discussion from official websites to Social Networks where citizens are already debating and take advantage of enhanced policy intelligence services based on fresh and relevant data. In particular, this platform will enable a centralized posting of contents and micro-applications (termed "Policy Gadgets" or, coining a *portmanteau*, "Padgets") to many different Web 2.0 Social Media at the same time, followed by the collection of various types of users' interactions (e.g., views, likes/dislikes, ratings, comments) and by advanced processing and analysis of resulting data in order to provide effective decision support to policy makers.

The aim of this paper is to present a policy analysis framework for this purpose, developed drawing from theories and concepts belonging to different scientific disciplines. The proposed framework constitutes the theoretical foundation on which the Decision Support Component (DSC) of the PADGETS platform has been designed. The development of such a decision support tool represents a first attempt to provide policy makers with a set of tools that may be precious in tackling grand challenges discussed at the beginning of this section.

Concluding these introductory comments, the paper is structured in five sections. Section two provides a theoretical background to the work presented. Section three illustrates the policy analysis framework proposed. Section four explains how such framework has been integrated into the PADGETS platform. Finally, section five provides some conclusive remarks on the value proposition of the decision support system proposed as well as some directions for future research.

2 Theoretical Background

The present section offers a concise overview of the most relevant strands of literature in order to provide an adequate theoretical background to the framework proposed.

2.1 E-Governance

There is no doubt that e-Governance has become in recent years a widely debated and researched topic among scholars. Despite this remarkable trend, "Governance" is still a problematic concept on which no agreed definition and vision exists. This uncertainty seems to amplify when the prefix "e-" is put before, which implies the exploration of some questions surrounding the impact of the Internet on Governments and Public Administration. Indeed, in order to arrive at a working definition of e-Governance, it is paramount to delve into whether ICTs change, eliminate or modify existing aspects of Governance and/or they create new problems and challenges.

In the resulting *mare magnum*, some authors [3] focus attention on the fact that one view of e-Governance entails an intense nexus with e-Democracy, particularly in terms of consultation and its mechanisms. A different perspective is reported by [4], whose vision is more operational and pragmatic, since the focal point is the application of ICTs to deliver governmental services, exchange information, perform transactions and integrate various standalone systems and services. Furthermore, a pure "institutional" vision is provided by Misuraca [5], according to whom "Governance is the exercise of political, economic and administrative authority necessary to manage a nation's affairs. Governance is the process of decision-making and the process by which decisions are implemented (or not implemented)".

In general, striving to put dissimilar visions under a common roof, it becomes visible that Governance connotes far more than just rudimentary functioning of Government: Governance is what the Government does in the exercise of its management, power and policy. Consequently, migrating to the electronic world, the concept and practice of e-Governance further encompasses e-Government: according to [6], the e-Governance concept, in fact, covers three distinct, yet related fields of application, i.e., e-Administration, e-Government, e-Democracy.

Taking into account the quality of a country's Governance, it emerges that this concept reflects the degree to which its institutions and processes are transparent and accountable to the people and allow them to participate in decisions that affect their lives. According to OECD, good Governance has eight major characteristics or dimensions [7]: it is participatory, consensus oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive and follows the rule of law. At a European level, five principles underpinning good Governance have been outlined in the White Paper on European Governance [8]: openness, participation, accountability, effectiveness, coherence.

2.2 E-Participation

As can be seen from the above lists of criteria, the participative dimension plays a vital role in the perspective of good Governance, since the participation demonstrates considerable potential to change the broader interactions between citizens and Government, improving the overall quality of engagement and decision making whilst widening the involvement of all citizens [9].

From a knowledge management perspective, in participative deliberations valuable "tacit knowledge" possessed by stakeholders is transformed into "explicit knowledge" by means of the so called "externalization" [10]. This process paves the way for a

new model of democracy, which is termed “participatory democracy” [11], combining decision making by citizens’ elected representatives with citizens’ participation, with the latter not replacing but supporting and enhancing the former.

However, despite rosy expectations and fervent impulses coming from the scientific community, the way Government’s consultation currently works never satiate the appetite of policy makers, due to the presence of notable difficulties (e.g., low attractiveness of ICT tools, limited set of questions, enormous wave of textual comments that hits policymakers) for citizens’ input to have a clear impact, as pointed out by [12]. As a remedy, new mechanisms are required to enable a public decision process open, transparent and participative in which citizens’ contribution is a paramount ingredient characterized by a significant impact.

Along this trajectory, the rise of social computing has recently attracted significant interest. In particular, the increased capabilities for Internet users to create contents, coupled with the birth of Social Networks, which have encountered dramatic success in terms of take-up, have driven the development of more and more virtual spaces for the expression of political views, problems and needs, which may ideally symbolize modern *agorae*¹.

Since Web 2.0 applications are already being used in Government not only for soft issues (e.g., public relations, public service announcements) but also for core internal tasks (e.g., intelligence services, reviewing patents, support decision making) [13], it is highly desirable to proceed towards a systematic exploitation of the emerging Social Media by Government organizations in the processes of public policies formulation, aiming to enhance a frictionless e-Participation: by doing this, Governments make a step towards citizens rather than expecting the citizenry to move their content production activity onto the “official” spaces created for e-Participation [14].

2.3 Decision Support Systems in the Public Sector

All different kinds of organizations, business, public, and non-governmental alike, are becoming aware of soaring complexity in decision making situations [15] [16]. Such complexity, inextricably related to the intricacy of systems [17], can be addressed through decision support tools which can enhance the quality of the decision process².

However, in a bevy of situations, multidisciplinary teams, top-notch skilled resources and world class computer suites do not suffice to cope with actual problems: a further need concerns the sharing and “externalization” of tacit knowledge already existing in the society. In fact, collective intelligence emerges as a key ingredient of a “distributed problem-solving” system [18] whose output can significantly enrich the decision process traditionally carried out by experts: in

¹ The term *agora* (Greek: *αγορά*) (pl. *agorae*) indicates the open place of assembly in ancient Greek city-states (*poleis*). *Agora* is many times used as equivalent of the roman *forum* (pl. *fora*), i.e., the public space located in the middle of a Roman city.

² The organizational decision making has its roots in the seminal contributions of renowned mavens such as Simon, Cyert and March; for a comprehensive discussion of these issues, see [19].

accordance with this trend, politics is progressively moving towards higher public engagement and cooperation in decision making processes.

Even though, as underlined by [20], the number of solutions in the area regarding e-Democracy, e-Participation and related fields is increasing, support systems for decision making are, however, still used mainly in narrow professional circles and have not found their way to political decision makers or to the public. The challenge of successful implementation of DSSs in the public sector, with engagement over the whole spectrum of decision making, is still unmet. In particular, in order to enhance the quality and effectiveness of the decision through knowledge harvesting, simulation of future scenarios and structured comparison of alternatives, DSSs depend on the availability and accessibility of timely, relevant and accurate information [21], which frequently represents the scarce resource.

To sum up, the implementation of successful e-Governance programs, which heavily leverage on the participative dimension, cannot ignore the presence of DSSs, as computer-based systems that help decision makers confront ill-structured problems through direct interaction with data and analytical models, notwithstanding the access to privileged channels aimed to enable a fully-fledged engagement: the evidence that “it has become impossible to restrict knowledge and its movement to castes of specialists” [22] makes researchers aware that “crowd wisdom” is not merely a Web 2.0 catchy buzzword, but is instead a strategic model to attract an interested and motivated platoon of stakeholders.

3 Towards a Cross-Policy Decision Support Framework

In order to design a decision support component for the PADGETS platform we started from two key broad categories of elements. Firstly we adopted a policy maker’s perspective, meaning that we focused on the manifold needs of daily policy making. Secondly, we considered the potential novelties introduced by the project, that is to say the several innovative approaches the project promises to introduce in the relationship and interaction between policy makers and stakeholders³: from this angle, the *leitmotif* is the exploitation of many Social Media at the same time in a systematic and centrally managed manner.

Keeping in mind these two cornerstones, we developed the architecture of a service that aims at informing the policy maker’s decision process (i.e., a decision support tool) by effectively using the knowledge collected in Web 2.0 Social Media through the engagement with a plethora of stakeholders.

Furthermore, we identified some potential threats to the effectiveness of our tool and we considered them as designing principles. These potentially critical issues concern the vast fields with which policy makers have to deal, such as the cognitive problem of synthesizing the distributed knowledge collected from stakeholders in

³ We prefer the generic term “stakeholders” to “citizens” because we think that citizens are only the largest kind of stakeholders interested in interacting with policy makers, and that institutions, which cannot be reduced to their single individuals, can be interested too in the innovative ways of participatory policy making introduced by the project. Hence, actors such as, for instance, producers’ and consumers’ associations, political parties, trade unions, corporations and charities, could be encompassed under the label “stakeholders”.

many different environments in order to take decisions, and the intrinsic dynamics of public knowledge and opinion.

Starting from the rich variety of policy fields, we decided to develop a decision support tool capable to be as much as possible “generic” and “horizontal”, meaning that it should be easily and effectively employed for any kind of public policy. This was done, among other reasons, to enhance the appeal of the DSS in terms of commercialization, i.e., in order to be turned into a marketable product. As a matter of fact, the possibility to reach a wider pool of potential institutional adopters allows to benefit from economies of scope and scale, that contribute to lower the unit cost of service provision. Secondly, considering the issue of synthesizing the widespread information collected through many different Web 2.0 participatory tools provided by the project, we started by interacting with local policy makers in order to identify the support they expect from such kind of a tool. In particular, policy makers would like to receive answers to the following five questions which are relevant during each phase of a public policy lifecycle (agenda setting, policy analysis, formulation, implementation, monitoring):

1. Are stakeholders aware of the public policy?
2. Are stakeholders interested in the public policy?
3. What stakeholders think about the specific public policy solution that the policy maker has proposed? To what extent they accept it?
4. Which are the barriers to policy awareness and interest, and which are the barriers to changes in public opinion about the policy?
5. Which suggestions are coming from stakeholders?

The identified relevant questions allowed us to design a support tool capable of collecting stakeholders’ suggestions through text mining and opinion analysis in order to answer question 4 and 5, and to synthesize the most relevant information collected from the stakeholders concerning the policy proposal along three basic dimensions (awareness, interest, acceptance) in order to devise responses to the first three questions.

We conceived the awareness-interest-acceptance-opinion framework by taking into account the needs of public policy makers and at the same time drawing from relevant theoretical frameworks developed in the disciplines of innovation studies and political science. According to innovation research of Rogers [23], the diffusion of an innovation occurs, in fact, through a five-step process, which is a type of decision making: awareness, interest, evaluation, trial, and adoption. Furthermore, OECD [7] identifies three stages of on-line engagement: information (for increasing stakeholders’ awareness), consultation (providing opinions about the policy) and active engagement. Also, the concept of policy acceptance is well-recognized in political science as it allows to understand the coherence between the proposed public action and the systems of values present in society, a necessary precondition for a successful implementation of the policy. The concept of acceptance may be seen from a normative point of view, or from innovation point of view⁴.

Finally, in order to capture the dynamic nature of the information we collected, we designed a component of the decisions support tool which relies on a System

⁴ For an example of EU funded research project on policy acceptance see [24].

Dynamics model strongly based on the collected data in order to provide policy makers with near future forecasts of stakeholders' awareness, interest and acceptance.

Summarizing, from a policy maker's perspective the value proposition of the decision support tool we designed may be recapitulated as follows:

1. A methodological contribution related to information classification, since the tool provides a well-grounded conceptual framework aimed to classify and aggregate data stemming from social engagement in light of an increasing level of stakeholders' involvement (awareness, interest, acceptance).
2. A reduction of information complexity, given by a set of peculiar traits (e.g., data aggregation along multiple dimensions, cross-platform data analysis, data projection into the real world, simulation of phenomena evolution in the near future) leading to a well-framed synthesis of unstructured (and sometimes inadvertent) society's input which could be used in order to forecast possible impacts of policies in light of surfacing *vox populi*.
3. A support to emerging Governance models, since it enables new ways for collecting, organizing and delivering information at different authority levels, opening-up on-going Governance models by letting a wider audience to contribute to the political debate.

4 The Architecture of the PADGETS Decision Support Component

The PADGETS Decision Support Component (DSC) is the analytic engine processing and analyzing the results of the PADGETS Campaign⁵ in order to extract useful information for the policy maker. It is, in other words, the software components which prepares the information for supporting policy makers.

The DSC relies on information coming from the policy maker, from Social Media platforms, and from the Padget⁶, and consists of two main modules: the PADGETS Analytics and the PADGETS Simulation Model. Whilst the Analytics module aims at grouping and synthesizing raw information and at solving possible problems of statistical nature in collected data, the Simulation Model aspires to provide future scenarios of opinion change.

In the following paragraphs we sketch the basic working mechanisms of the DSC, concluding with a discussion on the effectiveness of the proposed design.

4.1 Inputs

The inputs of the DSC model come from three sources, i.e., Social Media, Padgets and policy makers.

⁵ In the project jargon, a PADGETS campaign entails a set of activities covering creation, distribution, interaction, monitoring and termination of one or more policy messages oriented towards a specific goal and related to the same theme.

⁶ In line with the project dictionary, a Padget is a resource (application or content), typically instantiating within a variety of Social Media platforms, which provides interactivity with stakeholders through an *ensemble* of native and "augmented" social functionalities.

Data coming from Social Media (retrieved by public APIs) and from the Padgets may be unstructured (i.e., open text content) or, otherwise, structured (i.e., users' actions and selections). Unstructured data flow into text mining activities which work inside the PADGETS Analytics module; structured data, for their part, constitute the inputs of quantitative analysis taking place in both PADGETS Analytics and PADGETS Simulation Model.

The inputs database contains two broad categories of information in terms of data organization:

1. policy maker data referring to the target stakeholders' group (socio-demographic data);
2. data stemming from social engagement, collected at the finest granularity (individual users' data) and structured according to the two dimensions of user and time (the user who acted and the time of action).

4.2 Basic Output Indicators

DSC outputs are developed along three different concepts in accordance with section 3: awareness, interest and acceptance.

From a Padget end user's perspective, each concept is a set containing the following ones, but not vice versa. Thus, a user interested in the policy must be also aware of the policy, but the opposite might not hold (i.e., an aware user can be not interested in the policy).

The distinction between the concepts is that acceptance concerns polarized judgments (i.e., positive and negative) collected by means of the Padget, interest regards all data generated by a proactive behavior by users in Social Media, and finally awareness is a matter of an only passive reception of the policy message in Social Media (i.e., without further spreading or commenting the Padget announcement owing to a lack of interest).

The typologies of outputs that it is possible to compute are three. Firstly, it is possible to draw the distribution of data over the main categories of stakeholders identifiable according to socio-demographic variables. Secondly, it is possible to project the data to the actual world. Lastly, estimates on how policy awareness-interest-acceptance will change in the near future can be computed through algorithms included in the Simulation Model.

4.3 Modeling and Simulation

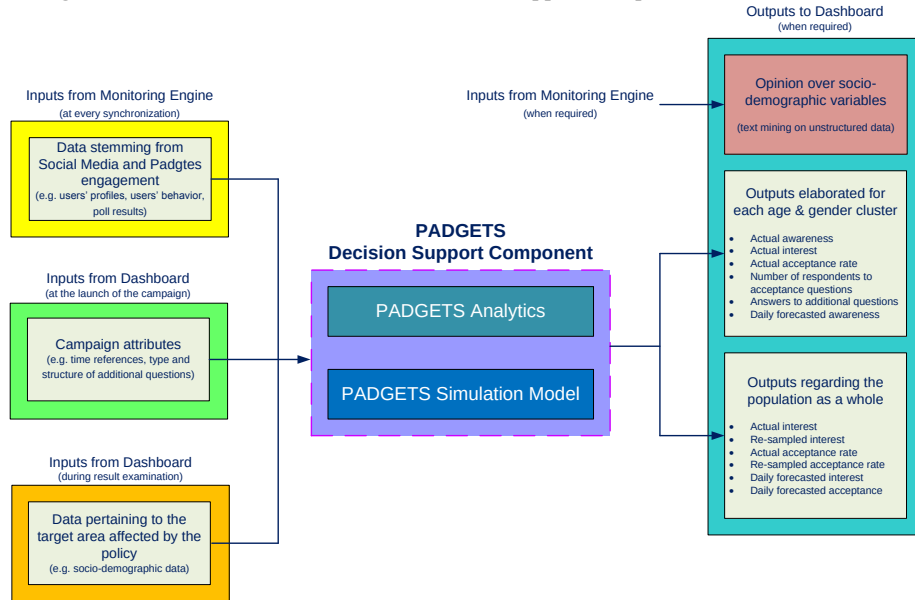
The PADGETS Decision Support Model structure (**Fig. 1**) shows how the two DSC modules transform the structured inputs coming from different sources in outputs useful for policy makers, while an external module carries out text mining and opinion analysis in order to determine stakeholders' suggestions. The figure, moreover, underlines that only the relevant information is presented to policy makers among the many we identified above.

Passing to the description of how the components work, actual distributions of awareness-interest-acceptance are obtained by simply grouping raw Social Media and Padget data according to socio-demographic variables.

In addition, in order to increase the real world significance of obtained results the re-sampling of raw data is computed: moving in this direction represents an attempt to remedy to possible underrepresentation of specific groups of stakeholders in the Social Media realm. To exemplify, older generations are likely to show lower penetration rates in Social Media: the resampling activity is thus aimed at reducing this bias in the estimation of current and future awareness, interest and acceptance rates.

Finally, awareness-interest-acceptance future scenarios are the most complex outcome. They start from the re-sampling of raw data, on which a procedure computes trends evolution along the dimensions of output concepts and of socio-demographic variables: a System Dynamics simulation is run in order to estimate the possible future outcomes and their probability of occurrence. Being the Simulation Model a System Dynamics one, there is the need for two main kinds of elements⁷, stocks and flows. Thus the Simulation Model is based on the identification of relevant stocks according to socio-demographic variables, and flows according to the above mentioned trends of evolution. In particular, the simulation is focused on how different socio-demographic clusters of stakeholders will change their level of awareness/interest/acceptance in the near future in light of intertwined social connections and resulting “viral” contagious phenomena; the rationale underlying the model entails that clusters are not independent, therefore several feedback loops and cascade effects can be at work testifying a blurred overlap of endogenous evolution and external influences.

Fig. 1. The structure of the PADGETS Decision Support Component.



⁷ For a complete introduction to System Dynamics, to its concepts and terminology, please refer to [25].

A final remark concerns the Simulation Model outcomes and their probability of occurrence. Given the stochastic nature of the simulation, the heterogeneity of collected data and the uncertainty on some parameters, the simulation will be run to explore all the possible outcomes of variations in parameters (including the random seed for stochastic processes); the resulting distribution of scenarios will be studied and presented to policy makers.

4.4 Effectiveness

The outputs of the DSC directly answer the first three policy makers' questions out of the five we want to address (section 3): presenting detailed basic and elaborated data about the three concepts of awareness, interest and acceptance, policy makers will obtain extensive answers to their questions, and they will also get an idea about future trends in society towards the policy on focus.

The remaining two policy makers' questions we plan to answer pertain to perceived barriers and suggestions coming from stakeholders. In this case the answers to policy makers are provided by the DSC as a whole, that is to say by all the results coming from the Analytics and the Simulation Model, on which we focused in this paper, and by analytic activities on unstructured data such as text mining.

In particular, the first three policy questions we considered as aims of the DSC are directly answered by singular groups of indicators provided by the DSC. For instance, if the policy maker is interested in evaluating the policy awareness in the population, the DSC provides a set of indicators related to awareness which describe the contemporary level of awareness and its near future trend. On the other side, the two remaining questions concerning emerging barriers and public suggestions are *per se* unforeseeable in their structure and content, and it thus becomes impossible to identify *ex ante* a structure to organize such information. For this purpose opinion mining [26] methods will be used. In this field, the effort is geared towards extracting opinions from unstructured human-authored texts (posts, comments *et similia*) having recourse to techniques such as feature-based sentiment analysis, topic identification and sentiment classification. Semantic analyses in this vein provide an insightful glimpse on "what people think" capable to conspicuously reinforce the governmental policy intelligence.

In conclusion, for the implementation of the DSC architecture we chose to rely on autonomous and platform independent software classes with data interfaces for communicating inputs and outputs with other parts of the project software architecture. Our choice has been to code this software entirely in Java (avoiding recourse to external libraries for System Dynamics modeling) in order to guarantee platform independence, eventual Web distribution and for relying on well-established libraries for the required activities of data management and regression.

5 Conclusions and Limitations

In this article we presented the preliminary results produced during the first year of activity of the PADGETS research project. In particular, the discussion focused on the policy analysis framework underpinning the DSC of the PADGETS platform. Such analytical framework was generated through a process of interdisciplinary integration

(mainly drawing from the diffusion of innovation and political science literature) as well as through a process of end users needs elicitation.

The intent behind the development of the analytical framework was to provide a first contribution towards the creation of a Decision Support System that could help policy makers in facing a number of relevant questions often arising through the policy cycle. This was done by introducing an innovation bringing together Social Networks and System Dynamics simulation. To date, in fact, the use of ICT tools for decision support has traditionally been a “closed door” activity usually carried out with static external inputs in the form of codified or unstructured data coming from different sources (e.g., statistical offices). Such approach presents a number of important limitations: the lack of a direct connection with the recent external reality on which the policy decision has to impact, an inherent delay present in the policy response due to the lead time necessary to collect and process the relevant data required for the analysis. To exemplify with a metaphor, such process could be compared to driving a car by only looking at the rear view mirror (an indirect and delayed input) rather than through the windscreen. The innovation brought by PADGETS consists in opening up the decision support process by integrating it with the activities carried out over Social Media Platforms. This allows to establish a direct link between the decision process and the external world as well as to reason on fresh and relevant information. This, once the necessary organizational processes are in place, should contribute to produce a much more responsive and effective style of decision making in Government. Going back to our metaphor, the innovation introduced by the Decisions Support Component of PADGETS aims at allowing decision makers to drive looking through the windscreen supported by an intelligent navigation systems able to anticipate some of the obstacles lying ahead (i.e., the predictive functionalities of the simulation module).

Finally, it is important to discuss also some of the limitations that characterize the solution presented, as they may represent an interesting starting point for future research. The resampling activity used for the generalization of the results in terms of interest and acceptance, for example, contributes to decrease some of the biases inherent in Social Media usage (e.g., age distribution) but it is far from producing a statistically significant representation of society. In addition, the implementation of a meaningful cross-platform tracking systems still presents a number of challenges having to do with identity management.

Concluding, although far from being error free, it is our firm belief that the framework presented constitutes a significant step ahead in helping policy makers in dealing with the challenges arising from the complexity that more and more may be found in modern societies.

6 References

1. Bauman, Z.: Liquid Modernity. Polity Press, Cambridge (2000).
2. PADGETS Project, <http://www.padgets.eu>.
3. Marche, S., McNiven, J.D.: E-Government and E-Governance: The Future Isn't What It Used To Be. Canadian Journal of Administrative Sciences, 20(1), 74-86 (2003).
4. Bose, S., Rashel, M.R.: Implementing E-Governance Using OECD Model (Modified) and Gartner Model (Modified) upon Agriculture of Bangladesh. In: 10th International

- Conference on Computer and Information Technology (ICCIT 2007), pp. 1-5. Dhaka, Bangladesh (2007).
5. Misuraca, G.: Literature Review: Multi-level Governance and ICTs. EXPGOV Project Deliverable, European Union, IPTS (2010).
6. Ferro, E., Molinari, F.: Framing Web 2.0 in the Process of Public Sector Innovation: Going Down the Participation Ladder. *European Journal of ePractice*, 9, 20-34 (2010).
7. OECD: Citizens as Partners: Information, Consultation and Public Participation in Policy-Making. OECD Publishing (2001).
8. Commission of the European Communities: European Governance: a White Paper. COM (2001) 428 final. Brussels (2001).
9. European Commission: European eParticipation Summary Report. Brussels (2009).
10. Nonaka, I.: A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), 14-37 (1994).
11. Pateman, C.: *Participation and Democratic Theory*. University Press, Cambridge (1970).
12. Johnston, P.: Transforming Government's Policy-Making Processes. *eJournal of eDemocracy and Open Government*, 2(2), 162-169 (2010).
13. Osimo, D.: Web 2.0 in Government: Why and How. JRC Scientific and Technical Reports, European Commission, Joint Research Centre, Institute for Prospective Technological Studies (2008).
14. Charalabidis, Y., Gionis, G., Ferro, E., Loukis, E.: Towards a systematic exploitation of web 2.0 and simulation modeling tools in public policy process. In: Tambouris, E., Macintosh, A., Glassey, O. (eds.) *ePart 2010*. LNCS, vol. 6229, pp. 1-12. Springer (2010).
15. Sterman, J.D.: Learning in and about complex systems. *System Dynamics Review*, 10 (2-3), 291-330 (1994).
16. Courtney, J. F.: Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. *Decision Support Systems*, 31(1), 17-38 (2001).
17. Beers, P.J., Boshuizen, H.P.A.E., Kirschner, P.A., Van den Bossche, P.: Decision-support and Complexity in Decision Making. In: 5th Junior Researchers of EARLI (JURE) Conference. Amsterdam, The Netherlands (2002).
18. Brabham, D.C.: Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1), 75-90 (2008).
19. Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C.: Past, present, and future of decision support technology. *Decision Support Systems*, 33(2), 111-126 (2002).
20. Benčina, J.: Web-based Decision Support System for the Public Sector Comprising Linguistic Variables. *Informatica*, 31, 311-323 (2007).
21. Kamel, S.: Decision Support Systems and Strategic Public Sector Decision Making in Egypt. Paper No. 3, Working Paper Series in Information Systems for Public Sector Management, Institute for Development Policy and Management. Manchester, UK (1998).
22. Lévy, P.: *Collective intelligence: Mankind's emerging world in cyberspace*. Perseus Books, Cambridge, MA, USA (1997).
23. Rogers, E. M.: *Diffusion of Innovations*. Free Press, New York (1983).
24. European Commission: Population Policy Acceptance Study – The Viewpoint of Citizens and Policy Actors Regarding the Management of Population Related Change. Final report, EU Research on Social Sciences and Humanities. Brussels (2006).
25. Forrester, J. W.: *Industrial Dynamics*. The MIT Press, Cambridge, MA, USA (1961).
26. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135 (2008).

6. Balancing system expressivity and user cognitive load in semantically enhanced policy modelling

Balancing system expressivity and user cognitive load in semantically enhanced policy modelling

Christos Tsarouchis, Declan O’Sullivan, David Lewis,

Knowledge & Data Engineering Group (KDEG),
School of Computer Science & Statistics,
Trinity College, Dublin, Ireland
{tsaroucc | Declan.OSullivan | Dave.Lewis}@cs.tcd.ie

Abstract. Policy Engineering is the process of authoring management policies, detecting and resolving policy conflicts and revising existing policies to accommodate changing resources, business goals and business processes. Policy authoring involves developing a policy rule base populated with policies that specify where actions specified on behalf of subjects may or should be performed on targets (resources). Policy engineering can be a daunting process in terms of complexity of the subjects, targets and actions used in rules, and the potential for conflicting policy rules to be committed to the rule base. Semantic modelling of rule elements can help automate the detection of such conflicts, but the additional layers of abstractions may themselves add to the complexity faced by the policy author. In this paper, we aim to assist the policy author by increasing the system expressivity with semantics, while at the same time minimizing the perceived cognitive load due to additional model complexity. We build on work with abstractions aimed at achieving this goal in the modelling of organisational grouping for subjects of policy-rules, which supports specific authority-based group abstractions to ease the maintenance of the subject model in the face of frequent organisational change. In this paper, we study this approach as used in combination with description logic based modelling of target semantics plus logic programming assertions across subject, targets and actions. This is performed through detailed analysis for policy authoring deliberations observed in a user evaluation of these modelling techniques.

Keywords: policy engineering, semantic modelling, description logic semantics, policy rule semantics

1 Introduction

Policy Engineering is the process of authoring policies, detecting and resolving policy conflicts and revising existing policies to accommodate changing resources, business goals and business processes [1]. Policy engineering can be a daunting process in terms of complexity of the subjects, targets and actions used in rules, and the potential for conflicting policy rules to be committed to the rule base. The policy author’s cognitive load [2] can be stressed in an environment of fluid organizational structure and changing resources, which are the targets for policy enforcement.

In this paper, we aim to assist the policy author by increasing the system expressivity with semantics, while at the same time minimizing the perceived cognitive load due to additional model complexity. Semantic modelling of rule elements can help automate the detection of system conflicts, but the additional layers of abstractions may themselves add to the complexity faced by the policy author. Other research initiatives have attempted to solve such problems by combining the efficiency and the expressive power of rule-based systems with the semantic richness found in Description Logics [3]. For instance, a rule-based system can express variables and causality relationships whereas ontologies on the other hand, can represent complex relationships between entities in a taxonomic manner that could simplify maintainability.

In this paper, for the modelling of policy subjects, the rule-based Community-based Policy management (CBPM) [4] will be used. This integration falls into the category of combining Description Logics (DL) [5] and Logic Programs (LP) [6].

The structure of this paper is as follows: for the remainder of section 1, the benefits of semantically enhanced policy subjects and targets, as well as the integration of CBPM with DL, and the resulting syntax will be presented. In section 2 a case study based on the semantic encoding of the Trinity College Statutes, together with some examples, will be shown. An intervention analysis as part of a user trial series is presented in section 2.2. Section 3 discusses the paper findings, while in section 4, related and future work will be presented.

1.1 Semantically enhanced policy subjects and targets

One of the requirements in policy engineering is the need to have an accurate and up-to-date view of the managed resources as well as of the actions that can be performed on these resources. This is particularly challenging in the case when not only these resources are frequently changing but also when their corresponding actions change as well.

This can be achieved by having an up-to-date snapshot of the resources and their corresponding actions in question into a structured, manageable and easy to query knowledge base (KB). This KB can be represented using ontological representations. This is because the use of ontologies can provide a centralized form of knowledge aggregation, when even complex relationships can be described and queried. Using OWL DL [7] for instance, the “equivalent” and “disjoint” relationships between resources and actions can be described. An example application that showcases such a system is presented in [8].

When modelling complex organizational structures, the flow of authority between subjects and targets needs to be modelled and preserved. Large organizations have decision-making groups, which are usually hierarchical in nature and often delegate the decision authority to other groups lower in the hierarchy. To model such concepts the use of the ontological class-subclass relationship is not adequate. For example, a company’s Director has authority over the Project Lead, but the latter is not a subclass of the former. In reality however, decision-making groups are more fluid and overlapping than the classic hierarchical organizational models that underpin role-based schemes. For this purpose, this research will use the Community-based Policy Management (CBPM) [4] framework to model policy subjects, which has been shown

to model the organizational aspects of policy subject models in a way that makes the resulting policy base more robust to organizational change [9].

CBPM uses the notion of Community as the primary grouping abstraction with the aim of allowing groups within the organization itself to define communities to naturally reflect the changing nature of decision making (i.e. policy setting) authority. Communities towards the top of the hierarchy have the wider membership and more general function, while those toward the bottom have more narrow membership and more specific function. The hierarchy is designed to support agility and autonomy, allowing new sub-communities to be formed and encouraging the delegation of decision making authority as far down the hierarchy as possible.

Using CBPM enables us to have a powerful permissions scheme in place. This scheme grants or denies permission to perform actions on resources and it can also be used to define obligations.

In this paper we therefore aim to investigate how different combinations of DL and LP impact on the overall system expressiveness and on the cognitive load placed upon the requirements engineer, while aiming for a similar level of policy engineering robustness for target models, which CBPM delivers for subject models.

We will now present in more detail, the inner workings of the integration between CBPM and OWL DL.

1.2 CBPM and OWL DL integration

The type of integration between CBPM and OWL DL presented in this paper, is of the hybrid type [3], which means that there is no creation of a new logical language. The resulting framework is called CBPM+DL. CBPM+DL is formed as follows: a hybrid knowledge base $K = (S, R)$ is a finite set S of DL axioms in the ontology language S and a finite set of hybrid rules R over R and S , including non-DL atoms. This definition follows the hybrid integration definition presented in [3].

Since the CBPM+DL method follows this hybrid approach, the semantics entailed in the hybrid language are derived from the semantics of its constituent components, namely CBPM and OWL DL. A hybrid rule r is obtained from the CBPM rule r' , by adding DL queries in the body. The latter are in fact additional constraints on the use of the rule r' , which have to be satisfied so that the rule r' is fired.

B. CBPM+DL syntax

The syntax of CBPM+DL is formed by extending the existing CBPM syntax with ontological predicates.

1) Community Membership Rule

The membership rule defines the restrictions, which need to be satisfied, so that a person P is a member of Community C .

Person(?P) $\wedge$ property.?P $\rightarrow$ memberOf(?P, Community)

Property is of type ontological object property restriction. This means that user-specific properties can be created.

2) *Authorization Rule*

Authorization is about granting or denying access (negative authorization) to a subject to perform an action on a resource. In this research we use the notion of resource authority, which is introduced in CBPM. A resource authority specifies what action on which resource, access is sought for.

The resource authority is defined as follows:

$\text{Action}(\text{?a}) \wedge \text{Resource}(\text{?r}) \rightarrow \text{ResourceAuthority}(\text{?ra})$

The resource authority is used as a predicate in the authorization rule:

$\text{comm}(\text{?comm}) \wedge \text{isa}(\text{?x}, \text{?y}) \rightarrow \text{posAuth}(\text{?comm}, \text{?ra})$

The $\text{isa}(\text{?x}, \text{?y})$ denotes the OWL is-a relationship, and in this case it means that x is a subclass of y . This rule can be read that as long as the “isa” relationship is satisfied, the community comm , is authorized to perform an action on a resource, as specified in the resource authority restriction.

3) *Implies authority rule*

The implies authority relationship describes the authority which is exercised by a Community, an Action, and a Resource on a Community', an Action' and a Resource' respectively, which are located lower in the hierarchy of authority.

$\text{Community ImpliesAuthorityOver Community'}$

$\text{Action ImpliesAuthorityOver Action'}$

$\text{Resource ImpliesAuthorityOver Resource'}$

' : lower in the hierarchical tree

For example, the Community called Directors ImpliesAuthorityOver the Community' called Projects Leads. The Action Read All ImpliesAuthorityOver Read File. The Resource Servers ImpliesAuthorityOver Database Servers.

2 The TCD statutes case study

For the last 3 years the Trinity College Dublin (TCD) Statutes Review Working Party has been commissioned with the task of reviewing and rewriting the Trinity College Dublin Statutes. The main task for the Party is not to introduce new policies but to update and rewrite the existing ones. The changes have been deemed successful and the new Statutes are currently in force. This well documented, real-world problem of statutes definition and interpretation in order to minimize the number of conflicts poses a big challenge. We decided to study the combination of CBPM and OWL DL techniques in order to tackle the encoding of existing statutes and the detection of conflicts when enforcing new and existing statutes.

Some of the challenges this effort has, is to capture the complexity of the organizational structure of the College, with the policies and obligations pertaining to each College body.

For this purpose the use of the CBPM was chosen as a candidate to model policy subjects. Also, due to the reasoning power of OWL DL, static conflict analysis could be performed and this was one driving factor for using ontologies.

However, the main challenge that we faced was finding the right balance between the use of CBPM and OWL DL in terms of system expressivity without incurring a

big penalty on the user cognitive load. Since it is envisaged that such a system would be used not only by experts in policy management, but also by domain experts (i.e. experts in law specification in this case), the creation of easy to use, custom tools that support these encodings is considered of great importance.

Such a tool has already been developed and usability studies are being performed. In the remainder of this paper however, we will not focus on the tool support, but on the task of finding a working balance between the various encoding options, during the encoding of the existing statutes. For that purpose, a user trial has been conducted which did not involve the use of any particular tool. This was done with the intention that the user trial results will not be skewed by potential usability issues the tools might have.

Five carefully selected participants took part in the user trial, each with a strong background on either policy-based management, ontologies, or logic programming. This is because the users would be asked to encode statutes using such techniques. The user trial involved the following a training session introducing all encoding methods, where examples of already encoded statutes were presented, and the actual user statutes encoding session.

The whole trial was recorded and in the end the users were asked to fill in a questionnaire.

The users were presented with an “encoding path” which they should follow, in the form of a step-by-step encoding methodology. This was done so that all users irrespective of their background, will consider all encoding options and not jump straight to the one they feel more comfortable with. It was also done so that the users won’t be intimidated by the number of encoding options available to them, especially during their first encodings.

The encoding methodology given to the users was the following:



All users started with CBPM, and if it was not possible to encode with CBPM or they were not happy with the outcome, they moved to the next available encoding method, which is CBPM+DL in this case. The same happened until the users traversed all remaining encoding methods, namely OWL DL, OWL DL + LP, and LP.

During the training sessions, the users were informed that there is no single “golden” encoding, and as long as they adhered to the encoding guidelines, they were free to come up with an encoding that could differ with the one suggested by the policy expert.

The OWL DL + LP method was in essence the use of the SWRL language [10]. When selecting LP in this trial, the users had to use the syntax of a generic rule language such as Drools [11].

2.1 Example statutes encodings

The following examples were used in the user training session.

1st example

The following example was presented to the users during the training session. The task is to encode the following clause from the TCD statutes:

“The Chancellor shall be a member of the Caput of the Senate and shall preside at its meetings”

Following the encoding methodology, the user has to try to use CBPM first.

The user can search for nouns, which are subjects and as a result can be candidates for being CBPM communities. Likewise, the user can search for targets. For instance, this clause can be broken down as follows:

C NCC C

The **Chancellor** shall be a **member of** the **Caput of Senate**

A R

and shall **preside** at its **meetings**.

C:Community, A:Action, R:Resource, NCC: Native CBPM Construct

The “Chancellor” is the subject and therefore it is a Community. We don’t know whether the “Chancellor” has authority over the “Caput of the Senate”, so the latter cannot be encoded as a Resource. So “Caput of the Senate” is a Community. “Preside” is an Action and “meetings” is a Resource, probably of type “Caput of the Senate” since the clause describes the “Caput of the Senate meetings”. The term “Member of” found in this clause, is an indicator that the “Chancellor” can be a member of the “Caput of the Senate” Community, and therefore can be regarded as a Native CBPM Construct.

We are now ready to encode the first part of the clause, namely “The Chancellor shall be a member of the Caput of Senate”. We will use the CBPM membership rule to achieve this:

Person(?P) $\wedge$ memberOf(?P, Chancellor) $\rightarrow$ memberOf(?P, CaputOfSenate)

This clause can be read as follows: For Person P and P is a member of the Chancellor Community, it implies that P is a member of the CaputOfSenate Community.

We can now proceed to encode the second part of the clause, “(The Chancellor) shall preside at its meetings”. We can use a CBPM positive authorization rule:

First we define the resource authority, which does not mention anything about the subject:

Resource Authority:

Action(Preside) $\wedge$ Resource(meetings) $\rightarrow$ ResourceAuthority(presides_meetings)

The rule that uses the given resource authority is the following:

comm(Chancellor) $\wedge$ ResourceAuthority(presides_meetings) $\rightarrow$ cl.posAuth(Chancellor, presides_meetings)

This rule can be read as follows:

For the Chancellor Community and for the presides_meetings Resource Authority, it implies that the Chancellor is given positive authorization to preside in Senate's meetings.

The encoding of this clause is a good fit for CBPM, because both the membership rule and the authorization rule can be applied easily. The same clause can also be encoded with the other remaining methods in the encoding methodology. For instance, for the OWL DL encoding, the clause breakdown is shown below:

OC OP OC

The **Chancellor** shall be a **member of** the **Caput of Senate**

OP OC

and shall **preside** at its **meetings**.

OC:Ontological Class, OP:Ontological Property

Following the OWL DL modelling, "Chancellor", "Caput of Senate" and "meetings" can be ontological classes. "Member of" and "preside" can be ontological object properties.

Thus, so far we have encoded the same TCD statute clause using two encoding methods, CBPM and OWL DL. The encoding choice is simply a matter of personal preference. However, in some type of clauses, such as those that have conditionals, the encoding process is not obvious, as user trials have shown.

2nd example

This clause includes a conditional:

"The Chancellor shall act as head of the University on ceremonial occasions"

"On ceremonial occasions" is the conditional.

Following a similar breakdown as in the previous clauses, the "Chancellor" and the "University" can be encoded as Communities and "head of" can be an Action. The question that arises is how to encode the "ceremonial occasions".

Let's assume that the condition "Ceremonial occasions" is encoded as a Community. This means that a new Community called Ceremonial is created and will be utilized on Ceremonial occasions.

We can use the CBPM authorization rule to encode this clause. First we create the Resource Authority:

Action(heads) $\wedge$ Resource(University) $\rightarrow$ ResourceAuthority(heads_university)

The authorisation rule reads:

comm(Chancellor) $\wedge$ memberOf(Chancellor, Ceremonial) $\wedge$ ResourceAuthority(heads_university) $\rightarrow$ cl.posAuth(Chancellor, heads_university)

The meaning of this rule is that the Chancellor, who belongs to the Chancellor Community and is a member of the Ceremonial Community and for the given Resource Authority, it implies that the Chancellor is authorized to head the University.

The encoding of this rule has shown that by restricting membership into a new Community, such as the "Ceremonial" in this case, statute conditionals can be represented.

However, during the presentation of the same example in the training session, all users agreed that the "Ceremonial" Community is an odd concept, and did not feel natural. Therefore, they were urged to seek alternative methods to encode this clause.

Using CBPM+DL, this clause would be encoded as follows:

```
cl.comm(Chancellor)  $\wedge$  isa(Ceremonial, Event)  $\rightarrow$  cl.posAuth(Chancellor, heads_university)
```

Using such an encoding there is no need to create the “Ceremonial” community. Instead, using the “is-a” relationship, a more “user-friendly” Ceremonial Class is created, and we define the Ceremonial Class as a subclass of the Class Event.

This CBPM+DL encoding has thus proven to be an efficient encoding method for cases like this.

It is worth noting that using OWL DL to encode and infer conditionals is very hard, and all users avoided that.

3rd example

Although OWL DL cannot encode complex conditionals, there is one type of statute clause, where it was particularly useful.

“There shall be up to six Pro-Chancellors, who shall be members of the Senate ex officio”

This clause is hard to encode in CBPM. It would require the introduction of a meta-rule (a rule about a rule) that would count the number of existing Pro-Chancellors. The CBPM syntax which is available for this user trial, does not cater for the use of meta-rules.

On the other hand, this is a clause where the OWL max cardinality restriction is ideally suited for its encoding:

Pro-chancellors is max 6

4th example

The users had to encode the following clause:

“The Chancellor shall be the primary Visitor of the College and University, and in the event of a disagreement between him and the other Visitor the opinion of the Chancellor shall prevail”

One way of encoding this rule is to use a similar analysis as in the previous examples. However, a more holistic view of this clause reveals that it would be a good fit for the CBPM implies authority relationship.

In fact some users selected this method as their preferred one:

Chancellor ImpliesAuthorityOver Other Visitor.

The users have used the “implies authority” relationship to describe the relationship between the “Chancellor” and the “Other Visitor”. By doing so, they have ensured that under all circumstances, such as a disagreement between the “Chancellor” and the “Other Visitor”, the “Chancellor” always prevails. This encoding has saved the users from encoding every single predicate of the clause. For instance, the term disagreement is not encoded.

This encoding has proven than CBPM when used in such scenarios, can be very beneficial both in saving rule editing, as well as resulting in smaller encodings, which can be more easily understood by a third party.

The use of the “implies authority” relationship in this case, is an elegant solution of avoiding the painstaking task of having to analyse complex clauses.

2.2 User trials

For this research, to evaluate the user cognitive load caused by policy engineering tasks, a number of metrics are currently used:

- the user prior-knowledge of the encoding method in question.
- the time required to complete all tasks.
- an intervention analysis, which studies the type of interventions performed by the policy expert, so that the users complete all tasks successfully.

These metrics are an indirect indication of the user cognitive load, since they do not include the use of any type of biometric sensor. The metrics currently used were selected on the basis that they represent an easy and unobtrusive method of collecting data. In this paper, only part of the intervention analysis is shown.

2.2.1 Intervention analysis

There were two main actors when conducting the user evaluation experiments: the users –being experts in either ontologies, CBPM, or rules- were asked to perform a series of tasks and complete a questionnaire in the end, and the policy/domain expert who conducted the experiment. The intervention analysis studies the type of interventions performed by the policy/domain expert, during the experiment, as a means to assist users. These interventions could be hints or clarifications, about conceptualizing the problem, about the syntax used, about spotting policy conflicts, etc. The number and type of interventions can be an indication of the user cognitive load.

For a user trial of 5 users, which involved the encoding of TCD college statutes using a number of different encoding methods, there were 5 types of domain expert intervention.

I1: The user doesn't know what to do, and the domain expert gives a hint

I2: The user asks for clarification about the syntax and the domain expert gives a hint

I3: The user asks for clarification about conceptualizing the problem and the domain expert gives a hint

I4: The user doesn't ask for clarification, the domain expert intervenes

I5: The user doesn't ask for clarification, the domain expert doesn't intervene

The intervention made by the domain expert for every user and for each task is shown in figure 2. This graph shows that during the first task, most of the users didn't know what to do and as a result, the domain expert intervened a lot. This is in stark contrast with the 5th task, where the domain expert intervened the least and three users completed the encoding without any intervention. This has to do with the fact that by the 5th task, the users had up to a certain degree acclimatized themselves with all the different encoding methods.

The 1st task was about the application of the CBPM syntax, and as a result most users needed time to familiarize themselves with its syntax.

The encoding of the 2nd task was not an obvious selection, and the users had to decide on their favourite method. This is depicted by the fact that almost all users

needed assistance in order to conceptualize the problem, hence the Intervention type 3 is shown as the highest for this task.

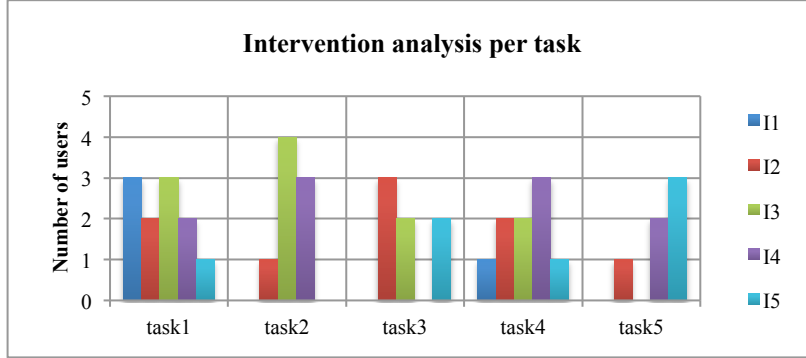


Figure 2: Intervention analysis per task

The 3rd task was about discovering that CBPM does not support meta-rules in its given syntax, and that users can use the OWL DL cardinality relationship. This is why the intervention type 2, which is related to clarifications about syntax, is shown as the highest.

The 4th task was about the encoding of a conditional. OWL cannot be used to encode and infer conditionals, but all other methods can be used. Most intervention types were used in this task, since it was the first time the users had to encode a conditional.

The 5th task builds on top of the 4th one, since again it describes a conditional, and that is why there are considerably less interventions needed compared to task 4.

The importance of user training in the encoding methods has been highlighted by the fact that the instances where the policy expert did not intervene to give hints, were increasing as users gained more experience in encoding.

3 Discussion

What has become obvious from analyzing the user responses, is that there is no one-size-fits-all encoding method, which doesn’t cause a rule clutter. Obviously, nearly every clause can be expressed in Logic Programs (LP) in the form of generic rules such as [11], but firstly, not all users are comfortable in using and interpreting LP rules and secondly the outcome being in the order of hundreds or thousands of rules, would be hard to maintain. Also, the benefit of having static rule conflict analysis that OWL DL provides, would be missing from an LP only case.

The 5-step encoding methodology used in the trial, does not necessarily describe a path from low expressivity to high expressivity, or from a given set of semantics to a richer set of semantics. This is because in some cases the semantics involved are just different, such as those between CBPM and OWL, and it is generally hard to agree on which one is semantically the richest.

What has been an interesting point of discussion is whether CBPM needs to be the first option for the encoding. This was done primarily because only after repeated use, could the users better utilize CBPM’s inherent principles. For instance, some of them used the CBPM “implies authority” relationship to describe a complex conditional.

Describing and inferring conditionals can be regarded as one of OWL DLs weakness, and thus alternative encoding methods must be selected.

Since the aim of the experiment was to showcase the strengths and weaknesses of each encoding method in a controlled environment, the syntax of each of the methods was explicitly defined and was available for the experiment participants to use. For CBPM, meta-rules were specifically omitted, since this is more akin to rule-based systems. Therefore, an alternative encoding method needs to be selected, when meta-rules are required, e.g. the use of OWL DL as shown in the example. Also, user-defined properties were left out of the CBPM syntax, since this is more tailored to OWL DL based encodings. The introduction of CBPM+DL encoding method can be seen as a bridge between two different worlds. Using CBPM+DL in the trials, the users were able to create custom properties and intuitive encodings in the cases where CBPM produced counter-intuitive ones, according to user responses. More on the user’s preferences regarding the various encoding methods can be found in [8].

4 Conclusions and future work

This research aims to assist policy experts and domain experts in the task of policy engineering, and especially in cases where the complexity of the subjects, targets and actions used in rules is high. The modelling of the Trinity College Dublin statutes poses such a challenge and for that purpose the use of semantically enhanced techniques has been suggested as a viable solution. This modelling involves the creation of a new encoding method called CBPM+DL, which uses CBPM to model organizational subjects and model the flow of authority between different parties. CBPM+DL also suggests the use of OWL DL to model changing policy resources (targets) mainly due to OWL’s ability to offer a taxonomical representation of complex resource relationships as well as static policy conflicts checking. This paper has shown that the use of technologies such as CBPM can be beneficial in terms of increasing the overall system expressivity, but unless systematic user training is performed, it is difficult to use in practice, due to increased user cognitive load.

Currently further trials are being performed, where the user’s prior knowledge in relevant technologies is correlated to the user’s performance in policy engineering tasks, such as policy conflict detection and resolution. Further evaluations are made to test the efficiency of a custom-built policy engineering tool, capable of modelling in the CBPM+DL form. This tool is based on the Graphical Modelling Framework [12], and its efficiency is compared against drools-based graphical editors [11].

Related work can be found in [13] where Rei, a policy framework that permits the specification, analysis and reasoning about declarative policies is presented. Rei adopts OWL Lite to specify policies and can reason over any domain knowledge expressed in either RDF or OWL. KAoS [14] defines basic ontologies for actions, actors, groups, places, resources and policies. Using an inference engine, KAoS can reason about policy disclosure, conflict detection, and domain structure. SWRL [10]

combines OWL with Horn-like rules. What differentiates the research presented in this paper from these approaches, is the use of Community modelling (CBPM), to model policy subjects, and their interactions, such as the delegation of authority between Communities. CBPM+DL also emphasizes in producing encodings, which are –in some cases- more intuitive than plain CBPM, as user trials have shown, thus assisting in the reduction of the user cognitive load. Since RuleML [15] is based on an XML mark-up language, it is tailored mostly for web use, rather than for the business level. SBVR [16] can be used in the future to model CBPM+DL concepts, such as Communities, the authority transfer between Communities, as well as the formation of Community Federations.

5 References

- [1] K. Feeney, C. Tsarouchis, and D. Lewis, "Policies as Signals in Collaborative Policy Engineering," presented at the PBAC, 2007.
- [2] D. Batra, "Cognitive complexity in data modeling: causes and recommendations," *Requirements Engineering*, vol. vol 12, pp. 231-244, 2007.
- [3] G. Antoniou, C. V. Damasio, B. Grosz, I. Horrocks, M. Kifer, J. Maluszynski, and P. F. Patel-Schneider, "Combining Rules and Ontologies. A survey," 2005.
- [4] K. Feeney, K. Quinn, D. Lewis, D. O'Sullivan, and V. Wade, "Relationship-Driven Policy Engineering for Autonomic Organisations," in *6th IEEE International Workshop on Policies for Distributed Systems (POLICY 05)*, 2005, pp. pp89 - 98.
- [5] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, *The Description Logic Handbook*, 2002.
- [6] C. Baral and M. Gelfond, " Logic programming and knowledge representation," *Journal of Logic Programming*, vol. 19/20, pp. 73-148, 1994.
- [7] OWL-overview. <http://www.w3.org/TR/owl-features/>.
- [8] C. Tsarouchis, D. O'Sullivan, and D. Lewis, "A study in the expressiveness of semantically different policy modelling schemes," presented at the 6th IFIP/IEEE International Workshop on Business-driven IT Management - BDIM, Dublin, Ireland, 2011.
- [9] R. Brennan, D. Lewis, J. Keeney, Z. Etzioni, K. Feeney, D. O'Sullivan, J. A. Lozano, and B. Jennings, "Policy-based integration of multiprovider digital home services," *Netw. Mag. of Global Internetwkg.*, vol. 23, pp. 50-56, 2009.
- [10] SWRL. <http://www.w3.org/Submission/SWRL/>.
- [11] Drools. <http://www.jboss.org/drools/>.
- [12] GMF. <http://www.eclipse.org/gmf/>.
- [13] L. Kagal, "Rei : A Policy Language for the Me-Centric Project," HP Labs 2002.
- [14] A. Uszok, et.al., "KAoS Policy and Domain Services: Toward a Description-Logic Approach to Policy Representation, Deconflictions, and Enforcement," in *IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, 2003, pp. pp 93-98.
- [15] RuleML. <http://www.ruleml.org/> - Scope.
- [16] SBVR. *Semantics of Business Vocabulary and Rules* <http://www.omg.org/spec/SBVR/>.

7. Semantics-powered Virtual Communities and Open Innovation for a Structured Deliberation Process

Semantics-powered Virtual Communities and Open Innovation for a Structured Deliberation Process

George Anadiotis¹, Panos Alexopoulos¹, Konstantinos Kafentzis¹, Efrosyni Konstantinou², Milan Stankovic³, and Manuel Waeckerle⁴

¹ IMC Technologies, Athens, Greece

{ganadiotis,palexopoulos,kkafentzis}@imc.com.gr

² City University, Centre for Performance at Work, London, UK

Efrosyni.Konstantinou.1@city.ac.uk

³ Hypios Research, Paris, France

milan.stankovic@hypios.com

⁴ Vienna University of Technology, Institute for Mathematical Methods in Economics, Vienna, Austria

manuel.waeckerle@econ.tuwien.ac.at

Abstract. In this paper we establish the connection between Virtual Communities, Social Media, Open Innovation and Semantics by presenting evidence to support our claim that these are the elements that can form the foundation for a new decision making process via deliberation, focused on expressing and matching Problems and Ideas for topics at hand. Furthermore, we outline the design of this process, based on existing work and well-known techniques and tools.

1 Virtual communities and Government2.0

The importance of connectedness, social networking ties and the ways that networks influence individuals have started to gain widespread attention and recognition [3]. This tendency has both catalyzed and been catalyzed by the advent of a new paradigm in computing, the so-called Social Media (SM). We hereby aim to show how SM empower Virtual Communities, how Semantics empower SM as knowledge elicitation platforms leading to Enterprise2.0 and Government2.0 and how we can leverage these phenomena via Open Innovation to build a structured deliberation process to lead beyond open, towards inclusive policy making [16].

1.1 Virtual communities as socio-technological phenomena

People increasingly connect with one another via ICT, a paradigm known as Social Media (SM) [1]. These SM fostered connections can be leveraged in order to build, sustain and utilize communities of interest on various domains, which can then in turn be leveraged in order to collectively perform tasks, an approach known as Crowdsourcing problem solving [10]. This can be attributed to two important indicators of social development we see on the rise -information literacy

and leisure- and represents a radical step in the evolution of the way communities form and operate, giving birth to the concept of virtual communities.

For the traditionalist, communities are closely linked and discussed in regards to notions of space, locality and neighbourhood, where physical co-presence of community members is seen as a defining factor. In this view, communities affected by technological change are frequently associated with disintegrated, fragmented, socially weak or isolated forms of social organisation. However, there is evidence that supports the persistence of communities and points towards an understanding of communities in relation to social networking, rather than space [8]. Under the post-modernist tradition that considers innovations in transport and technology to eliminate space through time, social networking can reflect meaningful and significant forms of interaction and participation [29].

Therefore virtual communities and social networks can play a significant part in shaping society and the economy, since they can reflect means of meaningful engagement, companionship and belonging. Lave and Wenger [13] talked about legitimate peripheral participation of agents in social communities, i.e. the process through which agents become members in social communities. They argued that, via learning, the meaning of the social world is constantly (re)negotiated among agents in communities of practice and, in turn, agents develop identities that are finely tuned with the constantly evolving meaning of the social world and progressively move from peripheral to central participation. In doing so, they continuously influence the existing system of relations within a community, while their identities are similarly influenced and eventually legitimate the participation of the agent in the community. Thereafter, shared understandings among legitimate members are again informed through interaction and new experiences of the world, community and the self. This shared understanding of the world enables the flow of knowledge between community members and helps them 'fill in' information and cultural references missing via SM user-generated content.

Even more importantly however, virtual communities and networks do not exist in isolation from the wider online environment and other communities within it. Incidents such as the Arab uprisings and the role that SM played in their organisation and expression highlight the potential of online media in reaching and connecting considerable parts of the world population. Such phenomena suggest that the processes of socialization, social change and reform, and the development of broader social dynamism via the extensive dissemination of knowledge in a format that is readily accessible to many, become shorter, immediate and significantly more inclusive. Communities create streams of political influence sustained via SM, counterbalancing the power of traditional media and shaping the political views and understanding of citizens across the world [4].

1.2 The role of Social Media and the Semantic Web in eliciting knowledge and supporting open governance: Government2.0

In today's globalised context a community need not be strictly defined by the legal, political and ideological framework of those elected to represent it [9]. SM and the enabling powers of the Semantic Web (SW) transfer the constitutional

powers of knowledge and information to the public in unprecedented proportions. This leads to improved civic equality and potentially enhanced civic action.

Recent research in multinational, knowledge intensive organisations has shown that the development of a critical approach will supersede issues of common concern. The ability to critically reflect, analyse and synthesize that is constitutive of a social self will be similarly applied at the level of online social interaction and engagement and manifest in political/civic choice and action [12]. That is, online content users and generators will use their critical thinking abilities more broadly in online and face-to-face interactions, public and private life and in relation to social, political and economic issues that attract their attention.

Post-WWII industrialist societies are typically characterized by rising GDP per capita, rising educational attainment, rising life expectancy as well as widespread adoption of reduced working times. For the first time, society forced onto an enormous number of its citizens the requirement to manage something they had never had to manage before: free time. This vast amount of free time, known as Cognitive Surplus, has up to now mostly been spent in consumerist activities and traditional one-way media, however with the advent of SM we see an increasing amount of this surplus being 'invested' therein - a development with tremendous implications from a knowledge elicitation perspective [23].

Even though tacit knowledge, or what the knower knows, is the most valuable distributed database in any organization, it is not easily accessible or explicit to those around us [11]. Traditionally tacit knowledge is exchanged by means of social interaction, but SM has enabled people to exchange tacit knowledge via ICT as "that's what social networking enables and that's what the exchange of tacit knowledge means" [11]. By doing so, socialization is essentially transformed to externalization and large scale knowledge-based communities may be supported. The advent of Web2.0 and its interlinking with the SW marks significant progress in this direction, giving birth to concepts such as Interlinked Online Information Societies [18], with applications in private and public organizations. The notion of Enterprise2.0, "the use of emergent social software platforms within companies, or between companies and their partners/customers" [14], has promptly been expanded to Government2.0: "the use of IT to socialize and commoditize government services, processes and data"

Government2.0 is a domain in its infancy, thus although it has received high calibre attention (e.g. [16]) many of the issues pertaining therein have been merely identified, not effectively dealt with. As stated in [28], "online engagement vastly increases the range, type and mix of expertise on which complex and challenging decision processes can draw. This is particularly useful where issues are complex, contentious or involve conflicting values and assumptions. Government2.0 allows access to more open, diverse knowledge base from which to draw to improve the chances of seizing an opportunity or solving a problem".

If "everybody is ignorant, just on different subjects", SM show the reverse also applies: everyone is knowledgeable, just on different subjects and SM provide unique ways to tap into that knowledge.

2 Open Innovation for Government2.0

Open Innovation [2], based on the notion of innovation economics, emerged as a new paradigm for Enterprises to develop novel technologies by seeking intelligent solutions and ideas beyond Enterprise borders. Soon, a number of websites started to emerge to facilitate the process of Open Innovation on the Web, serving as mediators for Enterprise R&D tasks. In this respect, we can consider the approach as part of the Enterprise2.0 paradigm, therefore we claim that a similar approach may also be applied to Government2.0, harvesting intelligent solutions and ideas beyond Government borders to develop novel policies.

2.1 Innovation Economics

Innovation economics got prominent with the seminal works of J.A. Schumpeter, who addressed the relationship of entrepreneurship, innovation and business cycle fluctuations. In his work [22], the entrepreneur is conceived as an energetic and active agent who attracts attention along deviant behaviour. This sort of behaviour does not rely on rational expectations - as per the homo oeconomicus doctrine - but originates from risk-taking under uncertainty. Successful innovation is seen as the enforcement of new combinations of existent things, like products or even production methods. This notion leads us to a concept of novelty under the aspect of emergence [32]. Novelty originates through the introduction of new knowledge into the economy by setting a new standard in the form of a socioeconomic institution [5], or generally in the form of a new information structure/ontology. New economic rules are created through the diffusion of this novelty on the basis of recombination, imitation and adaptation. Innovation is an emergent property of the diffusion of an invention, by that we understand emergence as the unintended establishment of a new quality on a higher level/layer as their original parts. The economic aspects of innovation can then be summarized as follows:

1. The creation and enforcement of new products or new qualities of products
2. The introduction of new methods of production
3. The creation of new industrial organizations
4. The disclosure of new sales markets
5. The disclosure of new sources of supply

Technological progress within organizational research and development represents a major address of the rather young discipline of evolutionary economics. Evolutionary change reflects a combination of gradual continuous systemic development, but features also discontinuous breaks in this development, which introduce something novel in the complex process of speciation. Technological progress can be considered in a similar way, where major innovations lead to the upcoming of really new socio-economic as well as cultural structures. The most influencing technological inventions go hand in hand with the rise of new economic epochs or a cultural Zeitgeist as shown in Figure 1.

Period	Date	Innovation	Saturation point
<i>First Industrial Revolution</i>	Circa 1800–1850	Cotton based technology; spinning weaving, etc.	1810 –end of Napoleonic Wars
<i>Second Industrial Revolution</i>	Circa 1850–1900	Age of steam; railways, shipping, heavy industry, iron and steel, etc.	1870s
<i>Third Industrial revolution</i>	1908–1947	Petrochemicals, internal combustion engine, electrification.	Inter-war slump 1920s and 30s
<i>Post-war Boom</i>	1947–1991	Consumer goods, electronics, etc.	1973
<i>Contemporary Era</i>	1991 – present	Internet, wireless technology, biotechnology, etc.	2010s

Fig. 1. Economic periods according to major inventions

The new literature on innovation systems originated from work on organizational routines within firms [15]. Organizational routines represent the key to successful innovation today; innovation got professionalised along design and establishment of innovation systems [6].

2.2 Problem formulation and expertise location in Open Innovation

Problem formulation is essential to achieving openness and diversity in the innovation process. The ability to speak to different communities of practice and reach people using different terminologies, common to their own field of expertise is identified as a major technological challenge for open innovation tools on the Web [25]. The appearance of problem solving marketplaces have marked a significant paradigm shift from the old "Define and try" to the new "Describe and search" approach [24]. The motivation behind the latter is the fact that ideas, and sometimes solutions and technologies to innovation problems might already exist, in form of online content, or in the minds of the Web users - experts in their fields. In order to leverage their potential for the purposes of industrial innovation, the first step is to identify and formulate a problem.

A major limitation is however the fact that no text can be equally appealing and understandable to all communities of practice, while texts that would be general enough, would at the same time probably be less useful. This limitation of human communication is hardly surmountable by writing a better problem text, however the use of metadata can help broadcast the problem to the right place and audience. The textual formulation of a problem is the most evident aspect of its communicative side, but other characteristics of the problem also play an important role in the problem dissemination and in the harvesting of solutions [24]. Principally we refer to the keywords, deadlines proposed for solving, challenge type, as well as the reward that is offered to the best solutions.

In order to create an ontology for representing problems, we analyzed the nature of information relevant for problems and innovation challenges. The main rationale is to enable actors to consume, and automatically treat the problem descriptions in a rich and meaningful way that would enable them to match the problems to the competencies of their users, or to the content that they may possess and that potentially contains solutions. An additional motivation was to

make explicit the nature of innovation challenge and the nature of rewards and IP transfer that is supposed to follow the acceptance of solutions. As a result we have produced an ontology of problem challenges - Problem Challenge Ontology (PCO), available at <http://ontologies.hypios.com/tag> , that responds to the needs of representing problem challenges for their meaningful exchange between innovation-related systems. A detailed description is beyond the scope of this paper, but let us note the main aspects of our ontology: Challenge Workflow and Status, Rewards, Transfer Type, Challenge Type, and Problem Composition.

In order to enable matching of problems to other content such as user profiles and user generated content, the ontology reuses existing ontologies such as FOAF, Tag Ontology and DublinCore. The topics of the problem are expressed using the existing property from the Dublin Core ontology `dcterms:subject` that is commonly used on the Web of Data. Furthermore, each problem may be assigned tags through the property `tag:tag` of the Tag Ontology which links problems to individuals of the class `tag:Tagging`. Topics and tags play a crucial role in matching the problems to potential solvers and solutions on the (Semantic) Web. We have evaluated the potential of the Linked Data cloud to serve the needs of expert finding in open innovation scenarios in [26]. Although a range of imperfections in the currently available data imposes the need for intermediary data cleaning and binding tools and heuristics, the current Linked Data cloud can serve those purposes rather well, while bringing more flexibility to the expert finding process compared to legacy expert search approaches. We have also built a flexible expert finding system based on Linked Data for the purposes of Open Innovation scenarios and raising problem awareness by targeting the expert in the problem domain and the fields related to the problem domain [27].

2.3 Idea Management in Open Innovation

Idea Management Systems evolved from mind maps & suggestion boxes, adding options to store, display and organize submitted ideas due to technological development. SM have further contributed to their evolution, extending submission boxes to idea capture methods, as richer and better organized user input data brought new opportunities for data presentation and selection [30].

On a practical level, exchanging and analysing ideas across different software tools and repositories is needed to implement the concepts of open innovation and holistic innovation management across different virtual communities. As a plethora of virtual communities keeps emerging, it is foreseen that despite the fragmentation there will be overlap in terms of domain of interest and ideas generated. In this light, a common framework for defining and exchanging Idea Management community output would greatly benefit Open Innovation.

Ontological frameworks proposed to provide a common language to foster interoperability between tools and to support the idea life cycle include the Idea Ontology [20] and the GI2MO Ontology [31]. Through the use of an ontology additional benefits like semantic reasoning and automatic analysis become available. By adopting an ontological framework to express content generated by

virtual communities, we enable expressing a variety of aspects about ideas in a structured way, thus enabling discovery.

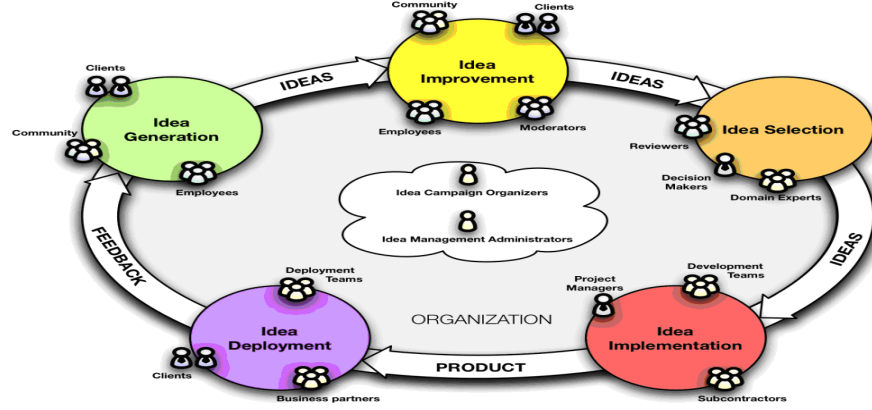


Fig. 2. Idea Management Lifecycle

We choose to adopt concepts, relationships and attributes that are particularly appropriate for describing such idea-related knowledge from the GI2MO schema, as it covers a broader scope and fits better to our purpose, especially in the way it facilitates categorization and tagging of ideas. The ontology captures a core idea concept that covers the 'heart of the idea' and further concepts to support collaborative idea development, including rating, discussing, tagging, and grouping ideas, as well as the idea lifecycle, which is of particular relevance when integrating idea management in the deliberation process.

GI2MO provides the property `hasTagging` which relates an idea with individuals from the class `tag:Tagging` of the Tag Ontology and the property `gi2mo:hasCategory` which does the same for individuals with individuals from the class `gi2mo:IdeaCategory`. The latter class has also the property `gi2mo:hasSubCategory` which is particularly useful for relating categories to each other and thus taking advantage of reasoners in order to infer multiple categories for an idea. Furthermore, there is also the property `sioc:topic` which relates resources such as forums, blog posts or news articles with ideas and thus compatible with the ontological representation in our community platform [7].

3 Towards a new deliberation paradigm: problems and ideas as part of a structured dialogue process

Having given background on virtual communities and the SM leverage on knowledge elicitation to support an open innovation process for Government2.0, we now turn our attention to defining and supporting a new deliberation process.

3.1 The need for a new deliberation process

Citizen abstention from political processes and growing indifference to public affairs and politics, even to the point of renouncing the cornerstone of democratic rights -voting in elections- is considered the greatest threat to democracy today. The EU Council Green Paper forecasts that abstention rates in Western Europe can reach 65% by 2020 [21]. Indifference and lack of trust in institutions reveal a "democratic deficit" which should be managed by political governance.

A promising approach to this direction is the use of ICT to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representatives. This approach, founded on eParticipation and eDemocracy but expanding their philosophy, is termed Citizen Engagement [19]. Citizen Engagement (CE) builds on both ICT and political science and bears the promise of reconnecting citizens and all other stakeholders to decision making and governance, by means of policy formulation and evaluation. All technical and legislative issues aside however, what remains the greatest challenge in today's CE landscape is actually achieving substantial CE levels [17]. Thus, an approach that takes into account today's socio-technological context is needed.

At its foundation, SM is a set of technologies and channels aimed at forming and enabling a potentially massive community of participants to productively collaborate. The Six core principles that underlie the value of SM solutions, and in combination serve as the defining characteristics that set SM apart from other forms of communication and collaboration are Participation, Collectivity, Transparency, Independence, Persistence and Emergence [1]. SM are becoming increasingly successful and widely used and at the same time more focused, a fact that can be seen as an oxymoron if we compare with the aforementioned CE bottleneck and consider how many these two domains seem to have in common. A way forward for CE could then be to learn from and adopt the SM paradigm.

3.2 Open Innovation as part of the deliberation process

In our existing work [7], we have taken a two-dimensional approach to facilitating CE via an eDeliberation process:

- a methodology has been developed, based on extensive study of related literature and deployed projects worldwide.
- an integrated SM platform has been implemented, providing appropriate tools to support every step and aspect of the methodology.

eDeliberation is a tight 'serial process' within a specific time-frame, with 7 concrete steps embedded in each deliberative cycle. Initially we have the agenda setting step, in which citizens may choose among a list of proposed issues for deliberation. After choice has been made, we move on to the discussion step, succeeded by a report publishing step, in which moderators summarize the discussion in a report. Then there is the voting step, in which citizens have to fill in questionnaires to quantify opinions and finally, after another report has been published to summarize the outcome of this step, an interactive online real-time

council takes place. The process concludes with the publishing of an overall report. We will now extend this approach in order to incorporate the aspects of Problem Formulation and Idea Management in the eDeliberation process.



Fig. 3. Methodology steps for eDeliberation

First, we introduce problem formulation in the process. After having chosen a general topic for the deliberation in the agenda setting phase, we break down this topic in specific problems to address. Problems can either be formulated using an online wizard-like tool that supports problem formulation in compliance with the problem ontology, or they can be harvested from the SW. This is in essence a matching task, so using the textual topic description that is the result of step 1 we can query semantic indices for matching problems. A number of problems will be the result of this process, therefore we need to be able to limit the number of problems to address in the deliberation. This can be done by factoring in a number of criteria: a) matching of problem to topic (automated metric) b) problem importance rating (user-generated aggregation) c) problem work group projected size (based on number of participants who join problem work groups) d) organisational preference, expressed via deliberation administrators. In order to support this process annotation, rating, group-creation and discussion mechanisms for each candidate problem are provided by appropriate tools.

The outcome of the problem formulation step will be a number of problems, therefore we introduce an idea management step in the process in order to collect ideas for dealing with each problem. Again, ideas can either be formulated using an online wizard-like tool that supports idea posting in compliance with the idea ontology, or they can be harvested from the SW. This is another matching task, so we need to match problem descriptions that are the result of step 2 to ideas either submitted by participants or found in other idea management systems.

The result of this process will again be a number of ideas for each problem, therefore we need to be able to limit the number of ideas to further develop in the deliberation. This can be done by factoring in a similar set of criteria as in the previous step, providing support to the process by a similar set of tools.

We have thus refined the general discussion step into discrete problem formulation and idea management steps, focusing the attention of participants to specific aspects of the open innovation process. A Proposal Shaping step has also been introduced in order to refine Ideas into well-rounded proposals for adoption by the hosting organization. Furthermore, by applying Social Technographics concepts certain user types (Creators, Conversationalists, Collectors etc) may

be identified and asked to contribute in the parts of the process where they best fit - e.g. Creators may be asked to submit ideas and Collectors to investigate Problems. Finally, the voting step on completed proposals gives participant input for the decision making that takes place in the Online Council step, determining proposal adoption, implementation and deployment. Thus with respect to the aforementioned Idea Management lifecycle, while the Implementation and Deployment phases take place internally in the organization, we insert an additional phase which we term Idea Scoping prior to Idea Generation. Idea Scoping consists of the Agenda Setting and Problem Formulation steps in the deliberation and represents what we believe is a necessary phase in open deliberation scenarios such as the ones conducted by Public Authorities, in which input on any given topic may be provided. Consequently, the process needs to be scoped and targeted carefully in order to be effective. The resulting deliberation process is shown in Figure 4 - omitting reporting steps for readability.



Fig. 4. Methodology steps for Open Innovation eDeliberation

3.3 Using Linked Data to share and match problems and ideas

There are 2 non mutually exclusive ways in which problems and ideas may be shared via the corresponding ontologies, both of which require a mapping from the underlying structure of the tools used (e.g. DB, CMS) to ontology concepts:

1. Direct sharing as Linked Data via SPARQL endpoint. Using RDB2RDF tools, an outbound Linked Data entry point to community content may be provided. Interested parties may then query endpoints directly via SPARQL.
2. Embedding RDFa in the generated web content. Using appropriate plugins, RDFa describing the content of generated web pages may be embedded in them, allowing for semantic indexers such as Sindice to harvest this information and make it accessible via their API.

Linked Data enable matching problems formulated with existing and new ideas, thus catalyzing problem solution and idea utilization. The enabling factor for this kind of matching is the tagging and categorization capabilities of the problem and idea ontologies. The availability of tags and categories in both problems and ideas allow for the semantic matching between the latter through the use of techniques ranging from simple SPARQL queries to methodologies such as Case Based Reasoning.

4 Conclusions and future work

Although this line of research is gaining traction and has a solid theoretical and technical anchoring, there is still gaps to be addressed before it becomes widely applicable. We identify the major gap to address not so much in terms of technical development, as the techniques we rely on have been successfully applied in a number of cases, but mostly in adoption in terms of organizational culture and governance. We remain hopeful however that the advantages of the Open Innovation paradigm and the ever more widespread adoption of Social Media will result in a change in organizational culture that fosters innovation and openness as the drivers for policy modelling and governance.

References

1. Anthony J. Bradley. *The Six Core Principles of Social-Media-Based Collaboration*. Gartner, December 2009.
2. H. W. Chesbrough. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press, 2003.
3. N. Christakis and J. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown and Company, 2009.
4. Couldry, N. and Curran, J. *Contesting media power: alternative media in a networked world*. Rowman & Littlefield Publishers, Inc., 2003.
5. Dopfer, K. , Foster, J., Potts, J. Micro-meso-macro. *Journal of Evolutionary Economics*, Vol. 14), pages 263–279, 2004.
6. C. Edquist. *Systems of Innovation: Technologies, Institutions and Organizations*. Routledge, 1997.
7. Anadiotis G., Alexopoulos P., Mpaslis K., Zosakis A., Kafentzis K., and Kotis K. Facilitating dialogue - using semantic web technology for eparticipation. In *In Proc. ESWC 2010*, Heraklion, Crete, Greece, 2010.
8. M. S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6), pages 1360–1380, 1973.
9. D. Harvey. *The condition of postmodernity*. Blackwell Publishers., 2001.
10. J. Howe. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business, 2008.
11. D. Jacobson. *Unlocking Tacit Knowledge with Social Networking*. PriceWaterhouseCoopers, 2009.
12. E. Konstantinou. *Knowledge Management in a global context: a critique of knowledge transfer and the role of knowledge workers*. PhD thesis, Stirling University, 2008.
13. J. Lave and E. Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.
14. McAfee, A. Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review*, 47(3):21–28, 2006.
15. R. Nelson and S. Winter. *An Evolutionary Theory of Economic Change*. Belknap Press, 1982.
16. OECD. *Focus on Citizens: Public Engagement for Better Policy and Services*. OECD, 2009.

17. E. Panopoulou, E. Tambouris, and K. Tarabanis. eParticipation initiatives in Europe: learning from practitioners. In *Proceedings of the 2nd IFIP WG 8.5 international conference on Electronic participation*, ePart'10, pages 54–65, 2010.
18. Passant, A., Samwald, M., Breslin, J., Decker, S. Federating Distributed Social Data to Build an Interlinked Online Information Society. *IEEE Intelligent Systems*, 24:44–48, 2009.
19. CROSSROAD Project. *Deliverable 2.1: State of the Art Analysis*. <http://crossroad.epu.ntua.gr/files/2010/04/CROSSROAD-D1.2-State-of-the-Art-Analysis-v1.00.pdf>, 2010.
20. C. Riedl, N. May, J. Finzen, S. Stathel, V. Kaufman, and H. Krcmar. An Idea Ontology for Innovation Management. *International Journal on Semantic Web and Information Systems*, 5(4), pages 1–18, 2009.
21. Schmitter, P., Trechsel, A. *The Future of Democracy in Europe : Trends, Analyses and Reforms. Strasbourg Green Paper on the future of democracy in Europe*.
22. J. A. Schumpeter. *Theorie der wirtschaftlichen Entwicklung*. Duncker & Humblot, 1997.
23. C. Shirky. *Cognitive Surplus: Creativity and Generosity in a Connected Age*. Penguin Press HC, 2010.
24. K. Speidel. Problem-description in open problem-solving. how to overcome cognitive and psychological roadblocks. *A Guide to Open Innovation and Crowdsourcing. Advice from Leading Experts*, 2011.
25. M. Stankovic. Open Innovation and Semantic Web: Problem Solver Search on Linked Data. In *In Proc. ISWC 2010*, Shanghai, China, 2010.
26. M. Stankovic, C. Wagner, J. Jovanovic, and P. Laublet. Looking for experts? what can linked data do for you? In *In Pre-proceedings of Linked Data on the Web 2010 (LDOW) workshop*, Raleigh, NC, USA, 2010.
27. Stankovic, M. and Jovanovic, J. and Laublet, P. Linked Data Metrics for Flexible Expert Search on the Open Web. In *Proc. ESWC2011*, Heraklion, Greece, 2011.
28. Government 2.0 Taskforce. *Engage: Getting on with Government 2.0*. <http://www.finance.gov.au/publications/gov20taskforcereport>, 2009.
29. Wellman, B. and Gulia, M. Virtual communities as communities: net surfers dont ride alone. *Communities in Cyberspace*, pages 165–172, 2003.
30. A. Westerski and C. Iglesias. The road from community ideas to organisational innovation: A life cycle survey of idea management systems. *A Special Issue of the Journal Web-Based Communities. Community-based Innovation: Designing Shared Spaces for Collaborative Creativity.*, 2011.
31. A. Westerski, C. Iglesias, and F. Tapia Rico. A model for integration and interlinking of idea management systems. In *4th Metadata and Semantics Research Conference (MTSR 2010)*, Alcal de Henares, Spain, 2010.
32. U. Witt. Propositions about novelty. *Journal of Economic Behavior & Organization*, Vol. 70, pages 311–320, 2009.

**8. Comelicopedia: a Collaborative, Semantically-enriched
Knowledge Base for the Sustainable Development in Mountain
Areas**

Comelicopedia: a Collaborative, Semantically-enriched Knowledge Base for the Sustainable Development in Mountain Areas

Alessio Gugliotta¹, Franco Alberti², Igor Jogan²

¹ INNOVA SpA, via Giacomo Peroni 386, 00131 Rome, Italy

² Regione Veneto, Direzione Urbanistica, Cannaregio 99, 30121 Venice, Italy

a.gugliotta@innova-eu.net

franco.alberti@regione.veneto.it

i.jogan@alice.it

Abstract. The paper reports on the expected achievements and the preliminary results of an ongoing work in developing a semantic wiki, supporting decision makers to effectively identify relevant knowledge about sustainable development in mountain areas. The underlying idea is to experiment a new, pragmatic approach in SD planning that could complement some of the existing and dominating tools that have been vigorously criticized by the community of both scholars and practitioners. The work represents a real-world example of adopting semantic and Web 2.0 technologies in complex domains, such as governance and policy making.

Keywords: Semantic Wiki, Sustainable Development, Policy Making, Use Case.

1 Introduction

Even though many years have passed since its first appearance on the international scene [1] the idea of Sustainable Development (SD) appears to be quite a contested matter [2]. According to the many international agreements signed with the purpose of addressing the problem of a more even development on the planet, SD is supposed to be a general umbrella under which policy decisions related to our life style patterns should be placed. Unfortunately this is not yet the case. On one side the latest appraisals produced on this matter depict a situation where most of the SD policy targets are far from being reached by European member states [3] while on the other scholars and practitioner are highlighting the need for a reviewed approach that would allow decision makers to access appropriate conceptual resources and institutional tools for devising sustainable solutions, being the ones available now under severe criticism. In particular, in the last decade, most of the efforts of regional and local authorities committed to SD have concentrated to enforce an evaluation procedure - the so called Strategic Environmental Assessment (SEA) - which has been in recent years and, in the light of this purpose, profoundly reconsidered [4].

The principal causes that are responsible for too little progress been achieved in this domain are those related to: (i) the high degree of approximation in defining the conceptual framework that inspires policy action [5]; (ii) the disproportionate limits in

either in know-how or know-way knowledge that affect public decision makers [6,7] and, on the other side; (iii) the lack in capacity shown by the experts and citizens in transmitting the available suggestions (objectives, ideas, projects, methods) to the top of the society where they can be collected and taken into consideration [8]. These conditions and related challenges are of particular interest in mountain areas because of their uncontrollable decline which is commonly ascribed, among other reasons, to a situation where local development occurs mostly on behalf of the inhabitants of the lowlands: the mountain population cultural behavior has been lining up with the exogenous cultural models and thus the losses in these areas are not only due to the environmental impoverishment but also to the threads that affect the local identities.

To a greater extent the aims of the SD can be pursued through the best use of the existing knowledge - being this either of conceptual or of practical nature - and the inception of a social learning process to which all the components of a local community have to participate. The authors attempt a new, pragmatic approach: to devise a method and a toolbox thanks to which decision makers can have effectively access to the needed knowledge that may allow them to free themselves of common conceptual schemes - that can bring them only to foregone decisions - and to make a jump towards sustainability. That is achieving decisions based on much broader, pertinent and updated knowledge coming from multiple sources and different layers of reasoning where full integration is obtained between different views.

2 Approach and Methodology

To achieve the aforementioned objectives, a semantic wiki [9] has been adopted as a baseline technology. In short, a semantic wiki combines benefits from two main waves in current Web technologies: social networks and semantic Web. The former enables involved stakeholders to share contents in a collaborative, bottom-up fashion. The latter provides the ability to capture, organise, evaluate and infer knowledge about the information brought by the pages, as well as the relationships between pages, in ways that can be effectively browsed and queried. Semantic wiki have being increasingly adopted in multiple domains with encouraging results. However, given the recognised complexity of policy making domains, the real challenge of our work is related to the elicitation, conceptualisation and formalisation of the underlying knowledge models (including the associated possible inferences), and the planning and implementation of the best approaches to feed, query and leverage the resulting knowledge base for the purposes described above. To assure qualitative and effective achievements, a staged, cooperative development methodology has been devised: *Kick-off*: the main classes and relations of the reference ontology are defined. A restricted group of domain experts actively participate to the generation, annotation and consultation of specific wiki contents. At this stage, the major sources of inspiration are, on one side, the planning theory, since the matter of SD is mostly concerned with planning tools and procedures, on the other, the textual analysis as the information that is collected is primarily made up of written texts. *Development*: After the first conceptualisation of the ontological model, the resulting ontology has been loaded into the semantic wiki to extensively generate instances through its tools. The work involves new domain experts and a selected group of possible end-users. The final aims of this phase are to (i) reach a high maturity level of the reference ontology and (ii) model a first set of SD querying functionalities. The querying interface holds

the responsibility of making the user aware of the issues at stake or better of the gaps that separates the real practices from other possible outcomes of ideal type policies. *Evolution*: a complete set of added-value functionalities, fully leveraging the reasoning capabilities and the available semantic annotation of the semantic wiki, are implemented and made available to an extended the set of possible users. At present, evolution is sought into two directions (i) the development of a geographic interface that can support the learning process also in spatial terms being territorial cohesion one of the prerequisites of SD and (ii) the extension of the system towards impact assessment functionalities that can enhance the appeal of our wiki also in a long time perspective. These functionalities represent the actual support for SD decision making and will be tested into the chosen use case to determine the actual potential and impact of the developed tool.

3 Use Case and Preliminary Results

The resulting semantic wiki, named Comelicopedia, has been devised in such a way as to semantically annotate and manage information from all the significant sources of knowledge, relevant to the issue of Comelico (a specific Italian mountain area) development. These sources are: (i) the policy documents issued by the responsible authorities at different scales (global, national, regional, local) (ii) the interviews to experts and community leaders (iii) a catalogue of the best practices from which decision makers can take inspiration for action and (iv) the comments that single citizens can fill in. The underlying ontology formalises the sources mentioned above, including: source structures, actors, geographical references and tags. Regarding source structures we identified two main elements of planning tools: context assessments and programme strategies. In terms of tags, we distinguish between two classes: semantic tags and keywords. The former are formalised a-priori by domain experts and are arranged in two distinct levels (top and specific tags), covering three main areas: sectors, themes and levels of innovation. The latter are introduced by users when semantic tags are not specific enough. The new defined tags can be then reused by all users when annotating other wiki contents. This drives the creation of a folksonomy that integrates the reference ontology [10]. To quote a few figures, Comelicopedia is made up of 603 annotated pages coming from 11 distinct knowledge sources; 76 semantic tags have defined by domain experts, whereas end-users have already introduced about 100 distinct keywords. According to SD principles, the querying system is devised with the aim of highlighting: sectors, levels and themes of policy decisions, the degree of integration between different actions and the intensity of innovation perceivable in each of the available resources. As a result, the users will interact in the decisional process in such a way as to: (i) Make them understand what are the challenges of this idea when applied to local development in mountain zones; (ii) Produce evidence about the way in which the idea of SD is perceived and possibly pursued at the different scales (multilevel decision making and coordination) and what are the 'visions' motivating different actors; (iii) Make citizens knowledgeable about the actions carried out by different stakeholders and their level of consistency, on one side, with the problems envisaged in the context analysis they subsume and, on the other, with the aims of the very concept of SD; (iv) Allow users to participate to the policy formation by contributing in terms of complementing the institutional decisions or even designing alternative

courses of action more or less supported to other practices (BP) or proposals that can be derived thanks to the knowledge management tools that are available in the wiki.

4 Conclusion and Future Work

Even though the semantic wiki that has been created needs a longer and systematic testing by the all possible users, the first testing shows that we are in front of an application that can really contribute to make the interaction of all stakeholders concerned in a policy modeling process more effective as long as all of them can share a common conceptual background. As compared to previously mentioned methods and procedures (i.e. SEA) that imply a periodical confrontation between decision makers and citizens based on a stepwise timing of the participation/negotiation processes, this technique allows for a much more intense interaction between the two institutional opponents where there is much more room for collective reflection and therefore learning. This allows decisions being shifted from the area of immediate interest linked merely to the pressure of the social demand to an area of a more aware and comprehensive knowledge. In this way, deliberation can aim at producing effects that are, at the moment the decision is made, less visible, but that can be appreciated in a longer span of time. Notice that once implemented, this approach would require a considerable revision of the existing institutional rules that govern public participation in development planning: that is a type of innovation that is by itself an important step toward SD. This is a demonstration that the method itself can contribute a lot to the achievement of this imperative social goal.

References

1. WCED - World Commission on Environment and Development (1987). *Our Common Future, Report of the World Commission on Environment and Development* (1987), Oxford: Oxford University Press. ISBN 0-19-282080-X
2. Connelly S. (2007), "Mapping Sustainable Development as a Contested Concept", in *Local Environment: The International Journal of Justice and Sustainability*, Volume 12, Issue 3, p 259 – 278.
3. Eurostat (2007), *Monitoring process towards a more sustainable Europe*, 2007 monitoring report of the EU sustainable development strategy.
4. Bina O., (2007) A critical review of the dominant lines of argumentation on the need for strategic environmental assessment, in *Environmental Impact Assessment Review*, v. 27. P5. 586-606.
5. Boothroyd P. (1995) *Policy assessment*. In: Vanclay F, Bronstein DA, editors. *Environmental and social impact assessment*. Chichester: John Wiley; 1995. p. 83–126.
6. Connelly S, Richardson T. (2005), Value-driven SEA: time for an environmental justice perspective? In *Environ Impact Assess Review*, v.25, p.391–409.
7. Owens S, Rayner T, Bina O. (2004), New agendas for appraisal: reflections on theory, practice and research. *Environment and Planning A*, v.36, p.1943–59.
8. Petts J. (2003), Barriers to deliberative participation in EIA: learning from waste policies, plans and projects. In *Journal of Environmental Assessment Policy and Management*; v. 5, p.269–93.
9. Semantic Wikis: A Comprehensive Introduction with Examples from the Health Sciences, Maged N. Kamel Boulos, *Journal of Emerging Technologies in Web Intelligence*, Vol. 1, No. 1, August 2009
10. Smith G. (2008), *Tagging: People-Powered Metadata for the Social Web*, New Riders, Berkeley, CA

Authors Index

Author	Paper
Alberti, Franco	8
Alexopoulos, Panos	7
Alonen, Miika	2
Anadiotis, George	7
Askounis, Dimitris	1
Boero, Riccardo	5
Borges, Marcos R. S.	3
Charalabidis, Yannis	1, 5
de Faria Cordeiro, Kelli	3
de Ramos Araújo, Lucas	4
de Souza, Jairo Francisco	4
Ferro, Enrico	5
Gugliotta, Alessio	8
Hotti, Virpi	2
Hypponen, Konstantin	2

Author	Paper
Jogan, Igor	8
Kafentzis, Konstantinos	7
Konstantinou, Efrosyni	7
Korhonen, Sami	2
Lampathaki, Fenareti	1
Lewis, David	6
Loukis, Euripidis	5
Machado Campos, Maria Luiza	3
Mendes, Pablo N.	4
Osella, Michele	5
O'Sullivan, Declan	6
Stankovic, Milan	7
Tsarouchis, Christos	6
Waeckerle, Manuel	7