



Transforming Social Media Data into Linked Enterprise Government Data Asset for Tracking Public Policy and Services

Mohammad Waqar
mohammad.waqar [at] deri.org

© Copyright 2011 Digital Enterprise Research Institute. All rights reserved.



Enabling networked knowledge

- How to cope with massive amount of real-time data generated on Social media by citizens?

- How can governments effectively harness the collective intelligence of citizens embodied in social media contents for different purposes?
 - Guiding decision-making process.
 - Sensing and tracking public sentiments on public policies
 - Monitoring quality of public services to improve QoS
 - Exploring different views regarding public policies, procedures and services

- Pulling social media content is challenging
 - Many sources such as Twitter, Facebook, Google+
 - Different APIs
 - APIs call limitations
 - Different return APIs formats JSON, XML etc.
 - Privacy Limitation
 - Noise in data

- Integrating social media contents into existing e-Government enterprise data requires significant engineering effort.

Employing:

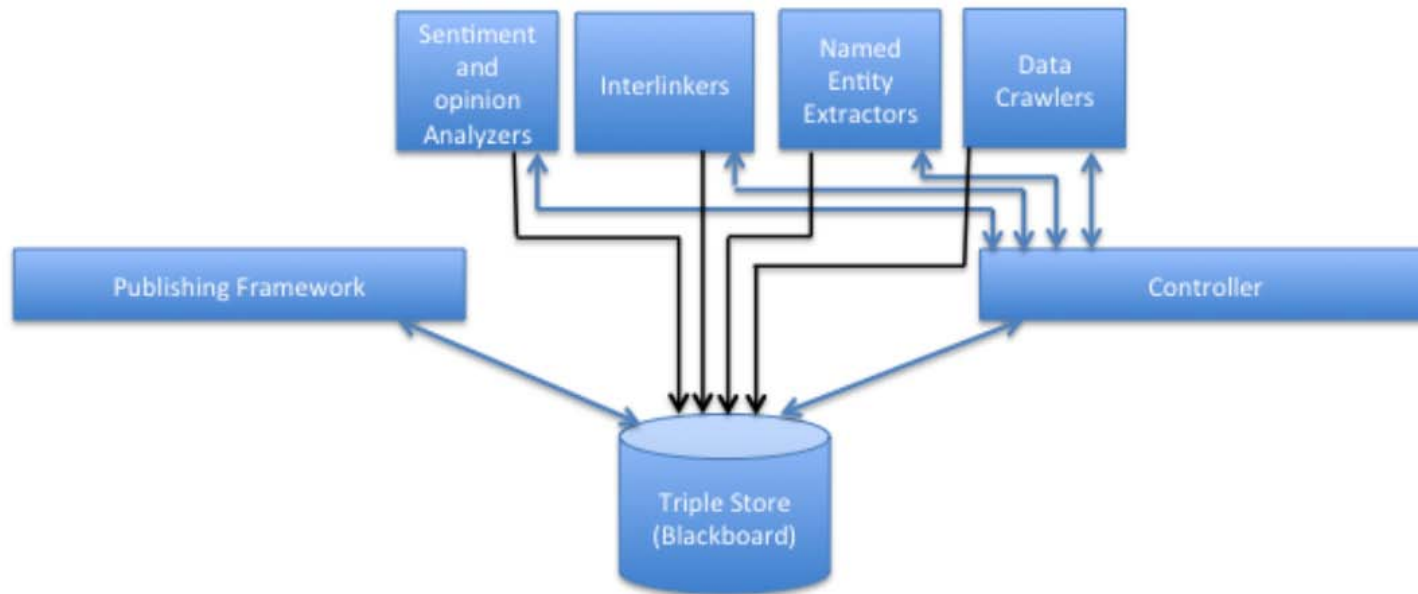
- 1) Semantic Web and Linked Data for information integration within the government enterprise
- 2) Natural Language Processing for Social Media Content or Text Analysis

To create: Social Media Linked Data Space

- Repository for holding social media content from different sources
- Transforms all captured contents into single machine readable format - RDF
- Enriches and integrates homogenized social data with existing Enterprise Data Hub

The design of the Social Media Linked Data Space is based on two popular design patterns

- The implicit invocation pattern.
- The blackboard pattern.

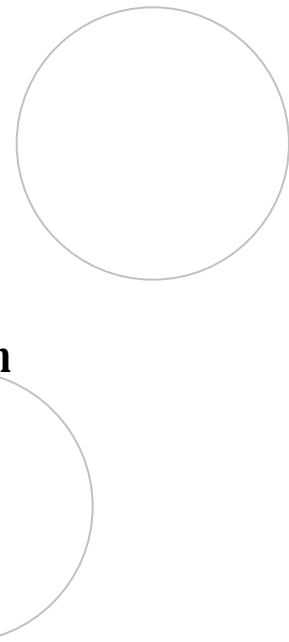


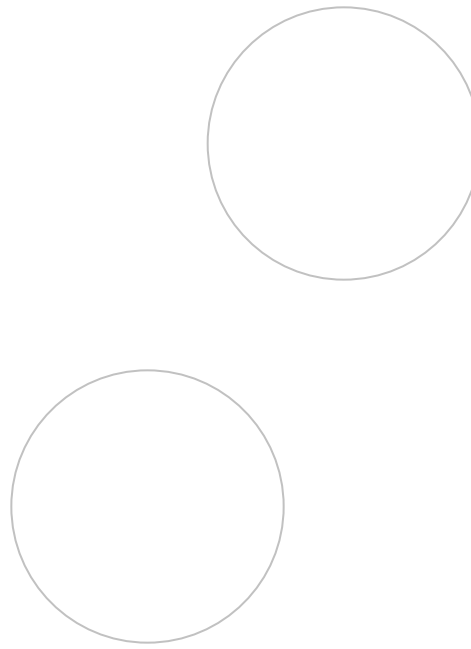
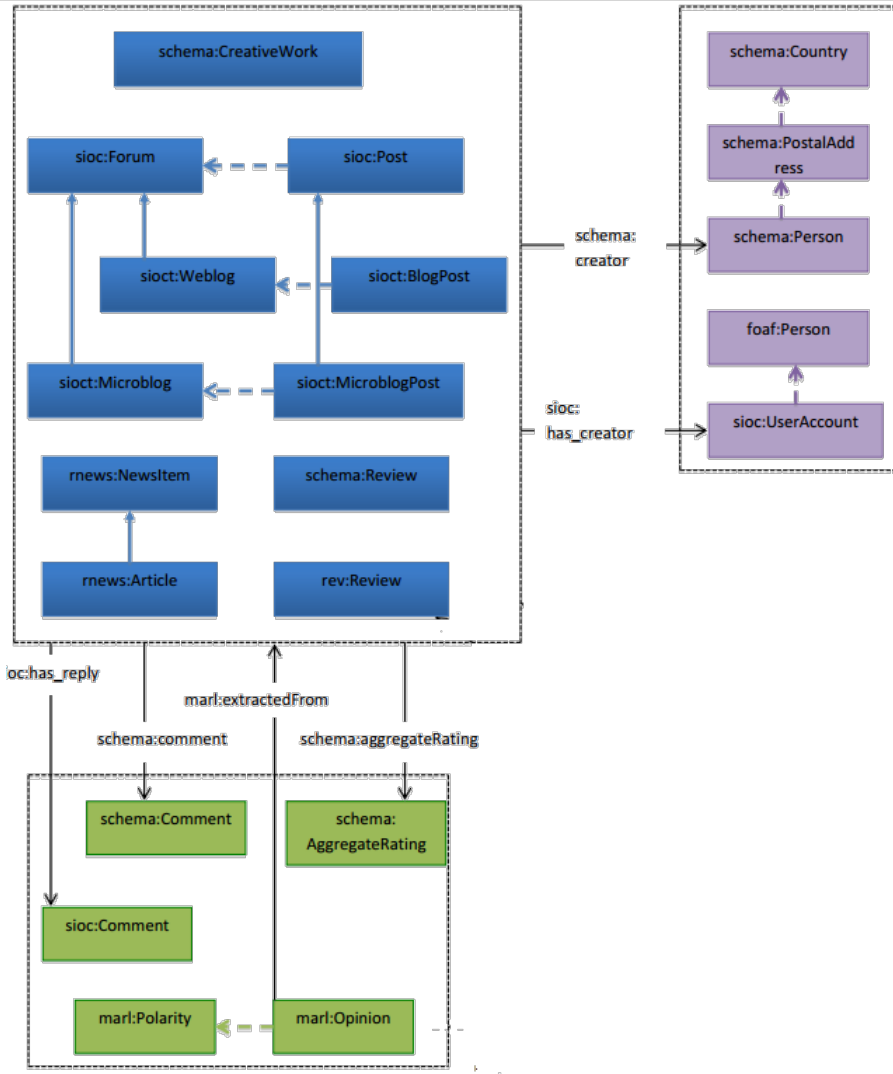
Technology Choice:

- One format for representation (RDF)
- One method for access the data (SPARQL)
- Linked to existing information hubs.

Design Choice:

- Simple and easy to scale.
- The components can be distributed.
- More components can easily be added to the system without disrupting the flow of the system.
- Uses standard protocols HTTP and SPARQL for communication.
- Platform and Language independent.



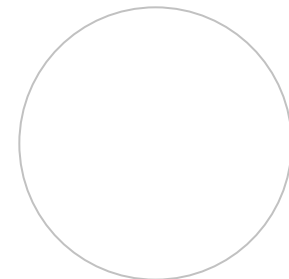
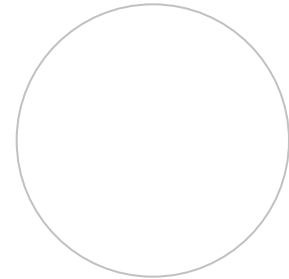


Techniques:

- Name Entity Recognition (NER).
- Interlinking

Tools

- Stanford NER
- DBpedia Spotlight
- Zemanta



Rich services can be built around Social Media Linked Data Space to improve decision making and service delivery, including the following:

- Tracking public policies and specific services differentiated by geographies and demography
- identifying ideological biases in public policy discourse
- Sensing public sentiments on on-going programs
- Detecting odd events for instance resulting from government-citizen communication gap
- Monitoring quality of services based on opinions

- 1) Proof of concept of Social Media Linked Data Space has been developed as part of an on-going project for product and service tracking.
- 2) Information on an initial set of entities have been extracted from 4 different social media platforms.
- 3) Plans to extract entities based on public policy and service vocabularies.

- 1) Philipp Frischmuth, Jakub Klímek, Sören Auer, Sebastian Tramp, Jörg Unbehauen, Kai Holzweißig, and Carl-Martin Marquardt, *Linked Data in Enterprise Information Integration, Semantic Web*, IOS Press, 2012
- 2) Evangelos Kalampokis, Michael Hausenblas, Konstantinos Tarabanis, *Combining Social and Government Open Data for Participatory Decision-Making*. In *Proceedings of third international conference on eParticipation (ePart 2011)*, Springer-Verlang, 2011,
- 3) Adegboyega Ojo, Elsa Esteveza, and Tomasz Janowski, *Semantic interoperability architecture for Governance 2.0*, *Information Polity* 15 (2010) 105–123 105, IOS Press,
- 4) 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), <http://ftp.jrc.es/EURdoc/JRC45269.pdf>



Thank you!

Acknowledgement:

Work is being carried out in the context of the Linked2Media FP7 Project

