

# Minutes of the W3C e-Government Interest Group Meeting

18 February 2013

## Present

Elsa Estevez (scribe)  
Tomasz Janowski (chair)  
Deirdre Lee  
Brand Niemann  
Adegboyega Ojo  
Lukasz Porwol  
Neeta Verma  
Mohammad Waqar

## Content

Tomasz: welcome and round of introductions

- Deirdre Lee: DERI, Ireland
- Mohammad Waqar: DERI, Ireland
- Elsa Estevez: UNU-IIST-EGOV
- Neeta Verma: NIC, India

Tomasz: the first speaker is Neeta Verma who will present about Open Government Data, India

Neeta:

- talk about OG implementation, challenges, and initiatives in India
- first formulated open sharing policy
- explained major stakeholders of the initiative
- policy provides the platform for open releasing government data
- releasing non-sensitive government data
- data should be released by all government agencies in central government
- all ministries and associated departments should ensure that all non-restrictive data should be released
- all agencies should nominate an officer - very senior officer leading the initiative
- all agencies should have focal persons
- agencies should have a plan for releasing data
- the officer will be responsible for receiving feedback from citizens and on releasing data
- some kind of standardization, metadata is necessary

- inter-ministerial platform was made available and they release guidelines related to the above issues through the platform
- each data controller should define a team
- human resources are needed to define what is high quality data
- people are also needed with visualization background
- also with knowledge in defining meta-data
- meta-data for the released data is available on the web
- they tried to develop an OG platform
- the project was completely developed on open-source
- the platform can be used by any country
- without major investments in ICT
- the platform enables connecting citizens with government and building a community
- the website is available at [www.data.gov.in](http://www.data.gov.in)
- citizens suggest the type of data they need and can make other types of suggestions
- it is possible to see how many data sets have been released by agencies, how many high-quality data-sets were released
- the website presents the list of all data-sets available, also data-sets can be found through the provider - which agency
- some statistics are also available - i.e. how many people saw the data-set, etc.
- they have developed a tool to convert the format of the data-sets provided by agencies
- agencies can provide the data-sets in different formats
- they have also integrated visualization tools for some data-sets; there are open-source tools
- slides 9 to 16 are examples of visualizations
- the visualization tools enables to view information with different levels of granularity
- they called the initiative "share-explore-experience"
- the tool enables to leverage citizen engagement - they can provide feedback on the quality of data, rank the relevance of data, they can also provide comments of their expectations about data-sets released and what they expect to be released
- the website provides a link to Facebook for visitors to easily share the data-sets with friends
- the feedback about interest in some datasets is pushed to the corresponding departments
- they are requesting feedback on the type of applications that are of interest to users
- the platform enables government to better connect to citizens
- the tool enables visitor to endorse datasets
- the main objective is to release high-quality datasets
- four steps: 1) tools and technologies for making data-sets available, for automating the whole process, including the generation of meta-data
- 2) motivation - talking to government departments, explaining that is free, and how to motivate citizens to use, and how to improve the usage of the data through visualization
- 3) capacity-building- explaining to agencies what is open data, what is high-quality data they are planning some kind of workshops to build capacity; there is a huge need for capacity-building

- 4) institutionalization of open data within government; there is a need of monitoring, integrating open data platform as part of e-Government platforms.
- any government application will need to release some kind of data as open data

Tomasz:

- many thanks to the speaker for great presentation and congratulations to the National Informatics Center on important achievements in the Open Government Data in India
- welcomes two more members of the group:
  - Brand Niemann, Semantic Community, US
  - Lukasz Porwol, DERI, Ireland
- opens the floor for questions

Tomasz: Besides access, does the policy cover the production of open government data?

Neeta: The policy covers both analogue and digital data. If the data is in analogue format, government needs to convert it.

Tomasz: What the process of moving data from closed to open formats?

Neeta: Releasing data refers to non-sensitive data. Government departments now understand what open data is and what data licensing means.

Tomasz: How do you measure high value and high quality? What are measurement metrics?

Neeta: The measurement is not precise yet. NIC is carrying out some research, but mainly they are requesting users to assess the relevance and quality of data.

Tomasz: The measures are subjective.

Neeta: Yes, they depend on the ratings provided by citizens.

Tomasz: How NIC ensures that open government data initiatives are sustainable, that they match the supply and demand for such data?

Neeta: Some datasets that are open now have been released periodically. One parameter to measure sustainability is the number of datasets that are periodically released by agencies.

Brand: What statistical services are provided to government agencies and how are they used in relation to social media in particular?

Neeta: India economic services and administrative/statistical services are two examples.

Tomasz: Thank you very much, Neeta, hope will join future discussions.

Tomasz: Welcomes Mohammed Waqar from DERI who will present how to transform social media data into linked data for tracking policy processes.

Mohammed:

- highlights the importance of social media and government
- social media empowers democracy
- due to its importance it will be a new media of communication between government and citizens in the future
- therefore governments allowed agencies to use social media
- there are some technical challenges
- related to scale and volume of data produced in social media
- how this data can be accessed efficiently
- how to track public sentiment related to public policies
- challenges related to getting data from social media
- there are backsides
- checking all social media is challenging and tricky
- some media provide APIs, others not
- the return data is in different format
- there are privacy limitations
- there may be some non-structured data
- requires significant amount of engineering efforts

Tomasz: Are there standard APIs for programmatic access to different social media platforms or do we have to invent special APIs for this?

Mohammed

- replies that there are no standards
- integrating natural language processing
- they integrated linked data repository to process social media data
- they sketched a high-level design for social media-linked data space
- he explains the technologies used - see slides 6 and 7
- all data is in RDF and can be accessed through SPARQL
- can use standard protocols for communications
- all models are consistent wrt semantics
- tools used are enumerated in slide 9
- they identified some applications (slide 10)
- allow to build services - like include decision making, tracking public services, sensing public sentiment, tracking public politics differentiated through geo-information
- Information on an initial set of entities have been extracted from 4 different social media platforms

Tomasz: Thank you for very interesting presentation and congratulations on important project.

Tomasz: Concerning the slide 11, you extracted entities from four different social media, can you explain what an entity is, and how extraction was done?

Mohammed: We extract different type of information - like info about products, services and locations

Tomasz: How?

Mohammed: Applying natural language processing.

Tomasz: Do you have sets of keywords related to public services and public policies and you search for these keywords?

Mohamed: Yes

Tomasz: welcome Gboyega. Adegboyega Ojo is leading the EGOV work at DERI.

Adegboyega:

- one of the key challenges is having standard vocabularies
- there are ongoing efforts to develop standard vocabularies for government
- so to facilitate information extraction from social media
- to move forward what is needed - availability of vocabularies
- ongoing efforts in terms of extraction they are using natural languages processing, which relies on heavy manual extraction tasks
- extraction from structured content relies on information from Wikipedia

Neeta: would like to connect with Adegboyega

Tomasz: delighted to see that W3C enables these types of connections

Tomasz: Is the platform planned to add more annotations or other structural information so extraction can be facilitated?

Adegboyega

- for extraction definitely ontologies are needed
- more specialized ontologies may be required to extract citizen opinions from social media platforms
- specialized ontologies will be required

Tomasz: Do you plan to include more organizational or policy recommendations to social media linked data space?

Adegboyega: Yes in two dimensions: government harvesting data and government linking data on social media, providing technologies to enable government to achieve these goals

Tomasz: Do you have to prescribe applications for social media linked data space or should the applications emerge naturally through the innovation process?

Adegboyega: We intend to provide a space where any application can be plugged to it.

Tomasz: Neeta, is the mandate to create open government data action plan by ministries extended to new initiatives which by design have to plan for open data?

Neeta: The policy requires maximizing open data. NIC is currently discussing how to automate the process of releasing data.

Tomasz: How is the demand assessed to maximize the use of high value data sets?

Neeta: There is a section on the open government data website for providing suggestions.

Tomasz:

- thanks both speakers – Neeta and Mohammed for excellent presentations, Adegboyega for answers, all for participating
- all are welcome to vote about topics to be discussed by the group in the future
- enumerates the current list of topics
- the highest ranked so far are interoperability, and transparency and accountability
- the poll is open until 1 March 2013
- the next meeting is scheduled for 18 March 2013
- since the open government data theme is concluding, we are looking for PhD students to help summarize, who would like to make use the presentations in their research
- thanks very much to Elsa for scribing
- very successful meeting across 12 time zones
- good day or night to all