IPHIX

Open Government Partnership – Australia

November 18, 2016

Public Consultation Submission – Open Government National Action Plan

Our company specializes in business data semantics and standards. We consult globally for Government digital transformation initiatives that cut red tape and reduce reporting burden, and provide software solutions for regulators and regulated entities alike.

Below are our comments on the "Australia's first Open Government National Action Plan – Draft for Consultation" currently accessible at https://ogpau.govspace.gov.au/files/2016/10/Australias-first-Open-Government-National-Action-Plan-Draft-for-consultation-Accessible.pdf.

Commitment 2.1: Release high-value datasets and enable data driven innovation

This commitment focuses on the identification and publication of valuable datasets, which is very important. Our suggestion is to give equal importance within this commitment to the usability of those datasets. Data standards play a key role in this.

The absence of application of data standards in open datasets is being flagged as a usability issue in several open government initiatives globally. Data standards enhance usability at different levels:

- Technical formats in which data is published. Formats that enable some level of integrity and validation of the data, such as XML and its domain-specific dialects, are preferable to formats that do not, such as CSV and similar text-based formats.
- Semantics and metadata, which make the information in the dataset machine searchable, processable, and understandable.
- In datasets where identification of parties and entities is important for usability, standards for consistent party identification should be considered. A further important distinction in this respect is between national identifiers,

IPHIX

800 W Broad Street #997 Falls Church VA 22312 • Ph. +1.202.379.4879 • Fax +1.571.723.4705 • www.iphix.net
EIN. 98-1198134

such as the ABN in Australia, and trans-national identifiers, such as the Legal Entity Identifier (LEI)[1]. The use of one type of identifier over the other affects the significance of data, for example for analysis across datasets from different countries. While ABN will continue as the dominant identifier of Australian business, the increasing globalisation of trade and a proliferation of identifiers makes the ideal of a single identifier difficult to achieve. Therefore, it will be imperative that solutions are implemented that allow cross-referencing of different identifiers at a minimum.

Commitment 3.3: Improve the discoverability and accessibility of government data and information

Data standards and semantics are equally important for this commitment. The focus in the document appears to be mainly on enhancements to technologies and platforms that enable users to more easily search and discover the data they need. We suggest that the primary focus should be on improving the quality and meaningfulness of datasets using data standards and metadata/semantics.

Resources spent on user interfaces, portals and other technical solutions to enhance search tend to have a short economic life and a sub-optimal return on investment, because they do not address data issues at the root, and focus only on how data is "served" to users. On the other end, discoverability of the information and efficient search are a natural consequence, and almost a by-product, when investments are targeted on data standardization and definition of appropriate metadata/semantics on published datasets.

While data.gov.au defines some metadata for the datasets, with the objective of enabling consistency of discovery of the datasets themselves, it does not establish guidelines for the discoverability of the data within the datasets. This level of discoverability is only attainable through the definition of a cross-government semantic registry.

In terms of formats for the publication of datasets, data.gov.au permits publication in any format, but CSV is the only format for which API access is automatically generated for the dataset. API access is obviously valuable, but the CSV format does not facilitate data integrity and has no validation capability. The result is that

---

[1] https://www.leiroc.org/lei.htm

IPHIX

800 W Broad Street #997 Falls Church VA 22312 • Ph. +1.202.379.4879 • Fax +1.571.723.4705 • www.iphix.net
EIN. 98-1198134

the use of the CSV format is encouraged, and conversely the use of more robust and standardized data formats ends up being discouraged.

We recommend that these considerations are taken into account in particular for milestones 2, 3 and 4, reproduced below for convenience.

"Milestone 2: PM&C to work with Data61 to create, circulate and gather feedback on design concepts and prototypes for improved search functionality and user experience on the data.gov.au platform.

Milestone 3: Finance to implement Phase One of the GrantConnect platform to enable public users to:

- register to receive notification on grant opportunities that match their self-defined profiles;

- watch forecast opportunities as they move from planning to requests for applications; and access grant guidelines for each opportunity and be notified about changes to grants processes.

Milestone 4: Finance to launch the pilot of the digital corporate and administrative reporting platform."

For example, Milestone 2 should, in our opinion, be reworded along the following lines:

"Milestone 2: PM&C to work with Data61 to create, circulate and gather feedback on:

- new publication policies and methodologies for data.gov.au datasets that establish a broader and deeper use of semantics in the context of a cross-government semantic registry, and of standardized data publication formats;

IPHIX

800 W Broad Street #997 Falls Church VA 22312 • Ph. +1.202.379.4879 • Fax +1.571.723.4705 • www.iphix.net
EIN. 98-1198134

- design concepts and prototypes for improved search functionality and user experience on the data.gov.au platform."

Kind regards,

Gianluca Garbellotto
CEO