

Mirek Sopek, sopek@makolab.com
LEI.INFO, www.lei.info
2153 S.E. Hawthorne Road, Suite 217
Gainesville, Florida 32641

July 5th, 2019

Re: Federal Data Strategy Draft Action Plan Comment (USBC-2019-0001)

Mr. Droegemeier, Ms. Kent, Ms. Potok, Ms. Kelley, and Ms. Roat:

I wish to thank you for this opportunity to provide written comments for the White House OMB Federal Data Strategy.

One of the most important aspects of data strategy is related to having a reliable digital identification mechanism. It is as important as data standards and data accessibility. It allows for the identification of each real-world entity the data is about, and for the identification of specific reference data about an entity. The question of identity is one that government will require to solve to maintain data relevance and credibility.

I'm co-authoring a book on the subject looking at the practicalities of current, technological market-leader solutions and I am also the founder of LEI.INFO¹, a Florida startup dedicated to using Legal Entity Identifier (LEI²) for rich data integration and new kinds of services. We were first to transform LEI data into semantically-enabled and ontology-driven, Open Linked Data formats and to protect it on Blockchain. We are now working on the conceptual enhancements that could help the LEI system become a truly foundational and infrastructural identification system for businesses on the global stage.

Within that remit I'm happy to attach a few very specific comments on the FDS action plan.

General Remarks

As noted above, we are convinced that the development and use of digital identification mechanisms is of critical importance for successful implementation of data strategy. We suggest an introduction of that aspect of data into Federal Data Strategy Year-1 Action Plan. This can be done at some specific Action Step or can be accomplished by coordinated activities at several steps. As we know, OMB already uses "OMB Control Numbers" and Action 7 assumes the use of them as identifiers. However, as the numbers were introduced in the context of the Paperwork Reduction Act, their scope is limited. What we refer to are identification mechanisms that are open-sourced and free to identify (and perhaps authenticate) the entities in the real world that are described by the data. When it comes to the class of entities that subsumes businesses and

¹ <https://lei.info>

² <https://www.gleif.org/>

other legal entities, including governmental organizations, it is suggested that the Legal Entity Identifier (LEI) is considered and evaluated as a candidate for general identifier.

Like any other system of global scope, the LEI system has both strengths and weaknesses. It's significant strength being its open-sourced implementation and the endorsement of G20, FSB and multiple regulators across the globe, while its main weakness is the slow pace of adoption related to some issues in its design and registration costs for legal entities. However, there are processes in place that will significantly reduce the burden of costs and improve its design toward the fully foundational and infrastructural character. We, at LEI.INFO, are actively working on the improvements.

In the following, action-specific remarks, I suggest more specific consideration of the digital identifiers for companies.

Action 3: Develop a Data Ethics Framework

The identification of real world entities in the data has essential impact on data ethics. In most situations this impact is positive and enhances data credibility. However, in some specific situations it may violate privacy or confidentiality of the entities being identified. There are multiple scenarios, including the concept of self-sovereign identity that could mitigate that kind of risk. We recommend the consideration and analysis of digital identification mechanisms into the development of a Data Ethics Framework.

Action 6: Pilot a One-stop Standard Research Application

It is highly advisable to include a consideration of the potential use of LEI data for the one-stop standard application for accessing federal data assets that are related to businesses or, in general, legal entities. The integration of LEI data into the application may come through the use of APIs, Linked Data patterns or bulk data repositories in semantically enhanced format.

The benefits for research and evidence building purposes are unquestionable – all activities that are in some way related to legal entities can use the LEI to link and integrate various data elements in an unambiguous way to provide consistent view about the entity under investigation.

Action 7: Pilot an Automated Inventory Tool for Data.gov

We believe that the design of the automated inventory tool planned for data.gov should consider the possibility of using Legal Entity Identifiers for the enhancement of the tool capacity to provide reliable and exhaustive information about the businesses and organizations identified by an LEI, which are available in the agency data inventories

As the tool already contemplates the use of the OMB Control Numbers as identifiers, the notion of digital identification is already present in the planning for the tool.

Action 8: Pilot Standard Data Catalogs for Data.gov

There is almost no doubt that the LEI system could be included in the government-wide data catalogue platform pilot. Its use for any data catalogue which contains data about any legal entities will greatly enhance the application capacity. The integration of LEI data into the application may come through the use of APIs, Linked Data patterns or bulk data repositories in semantically enhanced format.

Action 9: Improve Data Resources for AI Research and Development

It is recommended to use semantic data models and modern digital identification mechanisms in the improvements planned for the data resources. Data Coalition can share with OMB its experience related to building large scale data repositories with semantically enhanced data elements – ready for AI Research and Development. This experience includes both creation of data patterns and ontology building with the help of machine learning methods.

Action 11: Improve Geospatial Data Standards

It is interesting to consider the potential use of the latest achievements in linking digital identification mechanisms with trusted geospatial data. LEI.INFO is working with XYO Networks (<https://ml.ms/geo>) on the integration of the LEI system with geolocation system based on an innovative Blockchain system. LEI.INFO declares it is ready to share the relevant know-how and experience with interested people.

Action 15: Identify Data Needs to Answer Key Agency Questions

While this action is about the identification of the agencies' data needs, it is advisable to suggest the possibility of inclusion of the digital identification schemes, including the one offered by the LEI system into the possible key questions – whenever they are related to businesses and legal entities. This would provide an already established layer of data that, in most cases, may satisfy agency data needs on specific corporate data and create an instant solution to identification needs of, at least, a subset of corporations. In addition, this would provide a proof-of-concept scenario whereby access to the existing data can quickly highlight additional data needs and make a case for them.

Action 16: Identify Priority Datasets for Agency Open Data Plans

It is highly advisable to include the datasets used by the LEI system into priority datasets. As the LEI system is open-sourced and free, and as LEI.INFO has created its semantically enhanced version – there exists an opportunity to do so quickly and efficiently. LEI.INFO has proven experience in creating an integrative approach to critical data access and we can easily replicate or modify existing processes to create fast, viable solutions to agency data needs.

Dr Mirek Sopek
LEI.INFO CEO
sopek@makolab.com