

El proceso de publicación de los datos gubernamentales abiertos acerca de la clase de la Seguridad Social brasileña del Vocabulario Controlado del Gobierno Electrónico (VCGE)

CLÁUDIO JOSÉ SILVA RIBEIRO

Universidad Federal del Estado de Río de Janeiro, Brasil

DURVAL VIEIRA PEREIRA

Universidad Federal Fluminense, Brasil

INTRODUCCIÓN

Actualmente se invierten diversos recursos en la mejora de la comunicación entre el gobierno y los ciudadanos. La principal iniciativa que utiliza el Estado brasileño para efectuar esta interacción entre los ciudadanos y el poder público ha sido el Gobierno Electrónico. La base fue la “Propuesta de Política de Gobierno Electrónico” que surgió en el año 2000 a través de la idea de establecer un nuevo paradigma cultural de inclusión digital centrado en el ciudadano, con la reducción de costes unitarios, la mejora en la gestión y calidad de los servicios públicos, la transparencia y la simplificación de procesos.¹ Son fundamentales la inclusión social y la ampliación del empleo de la tecnología de la

información por el ciudadano, lo que permitiría la universalización y la democratización del acceso a los servicios.

Una de esas iniciativas se vuelve hacia los datos gubernamentales abiertos, lo cual puede considerarse como la “oferta de bases de datos en estado bruto, para que se manipulen, filtren o crucen con otras libremente, incluso permitiendo la construcción de nuevas aplicaciones y conocimientos por la propia sociedad civil.”² De esa manera, la publicación de los datos gubernamentales abiertos utiliza las tecnologías de la Web Semántica para que su empleo proporcione mejoras en la accesibilidad y reutilización de los datos.

Además, para que se alcance ese propósito, resulta imprescindible utilizar vocabularios para la descripción de los datos gubernamentales. Los vocabularios pueden utilizarse de dos maneras como esquemas: proveyendo un conjunto de propiedades que pueda tener un objeto, y como esquemas de codificación de vocabulario, describiendo el rango de valores que una determinada propiedad puede tomar formalmente. La utilización de vocabularios interconectados ayuda a adquirir conocimiento porque controla estrictamente y contextualiza los datos (conceptos, objetos y otros). Este enfoque permite trabajar con procesos similares de metadatos, y hace viable vincular vocabularios y registros de datos. Esto provee una infraestructura que permite el efectivo descubrimiento de la información y, por consiguiente, su empleo.³

La Empresa Tecnología e Informaciones de la Seguridad Social (Dataprev), de Brasil, destinó sus esfuerzos a la publicación de datos en el Ministerio de la Seguridad Social (MPS). El alcance del primer proyecto se dedicó a las informaciones sobre accidentes de trabajo, recogidos de los datos descritos en el *Anuario Estadístico de Accidentes de Trabajo* (AEAT). A través de estas informaciones se puede,

por ejemplo, construir un diagnóstico más preciso acerca de la epidemiología de estos accidentes y propiciar la elaboración de políticas más efectivas para las áreas relacionadas.⁴

Para que los principios de datos abiertos se cumplan completamente es necesaria la organización semántica de los datos publicados. Por eso, investigamos instrumentos de representación semántica de informaciones de seguridad social existentes y, por medio de esta investigación, localizamos el Vocabulario Controlado del Gobierno Electrónico (VCGE). Después de analizar la clase del asunto Seguridad Social dentro del dominio del Gobierno Electrónico brasileño encontramos que el dominio representado se podría revisar para ampliar el alcance de términos y de posibles usuarios.

De esa manera, este trabajo tiene por objetivo proponer una revisión de la clase de Seguridad Social representada en el VCGE, a fin de ampliar y corregir la representación que está disponible actualmente, con lo cual se abren posibles caminos al empleo de la semántica en la publicación de datos abiertos por Dataprev.

REFERENCIAS TEÓRICAS

Hoy, la mayoría de las informaciones difundidas por el gobierno está publicada en formatos propietarios o con tecnología que restringe su acceso. Sólo publicar los datos no los vuelven accesibles. Es necesario permitir desagregarlos, reprocesarlos y reutilizarlos, y buscar nuevas aplicaciones y utilidades.

Los datos gubernamentales abiertos constituyen la publicación de informaciones del sector público “en formato bruto abierto, de manera de hacerlos accesibles a todos y permitir su reutilización, como la creación de *mashups* de

datos G.”⁵ Además, los datos gubernamentales abiertos propician “el análisis de estos datos según varios puntos de vista; esta diseminación permite que la sociedad construya una visión más clara sobre el desempeño del gobierno frente a las metas establecidas, así como sobre el desarrollo de políticas públicas.”⁶

Los datos gubernamentales deben publicarse abiertamente por tres razones: para “aumentar la conciencia ciudadana sobre las funciones del gobierno para permitir mayor responsabilidad; [para] contribuir con informaciones valiosas sobre el mundo; y [para] permitir que el gobierno, el país y el mundo funcionen más eficientemente.”⁷

Al proveer significados a las “cosas” en *Web*, de modo que el lenguaje se descifre por máquina, capturando, procesando y relacionando datos de distintas fuentes, la *Web Semántica*⁸ hace viable que la publicación de los datos gubernamentales abiertos se encuentre por quien la reutilice, dándoles otra utilidad a éstos. De esa manera, la publicación de datos gubernamentales abiertos debe basarse en la *Web Semántica*, porque así los datos y las conexiones entre documentos recibirán significados definidos, con metadatos en un formato estandarizado, lo que permite la comprensión de personas y ordenadores.⁹

Adicionalmente, la Organización del Conocimiento contribuye en apoyar la estandarización lingüística necesaria en la *Web Semántica*. Puesto que este campo del conocimiento es responsable de importantes estudios sobre vocabularios estructurados y formalizados, estos instrumentos de representación del conocimiento pueden permitir que la *Web* trabaje con los conceptos y sus relaciones, para una efectiva recuperación y utilización de la información.

En las Ciencias de la Información, *taxonomía* es una de las clasificaciones que se basan en las jerarquías de asuntos,

actuando como instrumento capaz de organizar y recuperar las informaciones y garantizando el control del vocabulario específico del área.¹⁰ Permite que se establezcan estándares de alto nivel para la ordenación y la clasificación de información a través del empleo de mecanismos de herencia en la Web Semántica.¹¹

Otro instrumento de representación del conocimiento utilizado en la Web Semántica, la ontología, se presenta como “un vocabulario común para una comunidad que necesita compartir información en un determinado dominio. Incluye definiciones de conceptos básicos en el dominio y las relaciones entre éstos, de manera que se puedan interpretar por máquina.”¹² Algunos estudios que implican Web Semántica denominan *taxonomía* a una especie de “ontología sencilla.”¹³

De esa manera, los vocabularios controlados, entre los cuales destacaríamos a las taxonomías y las ontologías, pueden contribuir a que se establezcan relaciones semánticas comprensibles; en el caso de la Web Semántica, no solamente por hombres, sino también por agentes de softwares. Además, cooperan en la interoperabilidad, detección de inconsistencias, formalismo para la representación y reutilización del conocimiento de un determinado dominio.

Diagnosticamos una proximidad, casi dependiente, entre datos abiertos, Web Semántica e instrumentos de Organización del Conocimiento, como ontologías y taxonomías. En realidad, hay áreas del conocimiento que tratan temáticas en común, en las cuales la asociación provoca el desarrollo recíproco.

METODOLOGÍA

El recorte de este estudio contempla la taxonomía como el andamiaje de sustentación para el desarrollo de ontologías

El acceso a la información gubernamental...

en la Seguridad Social. Se trata de un estudio exploratorio que buscó identificar puntos en la clase de Seguridad Social del VCGE que podrían mejorarse. De esa manera, la metodología se dividió por etapas:

1. Selección de la clase de asunto sobre Seguridad Social del VCGE, en el dominio del Gobierno Electrónico brasileño, como recorte a observarse.
2. Identificación de posibles inconsistencias relacionadas con la categorización, especialmente sobre: el lenguaje elegido (natural o artificial); cobertura de los términos pertenecientes al alcance de la clase de asunto estudiado, y relaciones presentadas entre los términos.
3. Sugerencia de propuestas para revisión/modificación de la clase de asunto sobre Seguridad Social.

El caso aquí presentado es limitado e ilustrativo, no exhaustivo. Nuestro propósito es, en este momento, proponer la discusión sobre la utilización del VCGE como instrumento de organización del conocimiento para la publicación de datos gubernamentales abiertos.

ANÁLISIS Y RESULTADOS

El VCGE es una evolución de la Lista de Asuntos del Gobierno brasileño (LAG). Es un esquema a utilizarse en el elemento asunto / categoría del Patrón de Metadatos del Gobierno Electrónico Brasileño (e-PMG). Para los gerentes de sitios y portales gubernamentales, los términos retirados del VCGE volverán más directa la presentación de los servicios disponibles en una estructura de directorio que se basa en sus indexadores.

El proceso de publicación de los datos gubernamentales abiertos...

El vocabulario estudiado presenta las relaciones genéricas y específicas entre los términos que caracterizan una taxonomía. Aunque los documentos sobre la creación del VCGE afirman que se utilizaron los principios de un tesoro, el VCGE se acerca más a una taxonomía, porque posee una cantidad reducida de relaciones; no presenta, por ejemplo, las relaciones partitiva y asociativa, siempre presentes en un tesoro.¹⁴

Cuando analizamos la clase relativa a la Seguridad Social brasileña en el esquema del VCGE (*Figura 1*), reconocemos que la representación que se hizo de la respectiva área no corresponde a su alcance temático. Términos importantes para el entendimiento de la Seguridad Social no están incluidos en el vocabulario.

Figura 1
Seguridad Social en el VCGE

The image shows a screenshot of the Vocabulário Geral do Governo Eletrônico (VCGE) interface. On the left, a tree view displays the hierarchy of terms, with 'Previdência social' highlighted in red. The right side of the interface shows a detailed view of the selected term, with tabs for 'Informações', 'Exemplos', and 'Sugestões'. The 'Informações' tab is active, displaying the following information:

- Termo Autorizado (TA):** Previdência social
- URI:** <http://vocab.e.gov.br/2011/03/vcge#previd>
- Termo Geral (TG):** Pessoa, família e sociedade
- Termos Específicos (TE):**
 - Benefício assistencial ao idoso e ao deficiente
 - Benefício previdenciários
 - Pensões
 - Previdência básica
 - Previdência complementar
 - Previdência do regime estatutário
 - Previdência do servidor
 - Reabilitação profissional

Disponível em: <http://vocab.e.gov.br/2011/03/vcge#previdencia-social>

Un ejemplo de la falta de cobertura que se constató anteriormente es que, si un ciudadano necesita buscar por una especie de pensión de la cual se beneficia, no la encontrará representada, porque el término portugués *Pensão* no está clasificado por el VCGE. Se ausentan de la clase del VCGE sobre Seguridad Social las pensiones mantenidas, actualmente, por el Ministerio de la Seguridad Social.¹⁵

Otro punto de relevancia es que la Seguridad Social está subordinada a la clase de “Pessoa, família e sociedade”. Sin embargo, otras áreas, como “Saúde”, “Trabalho” y “Educação”, aunque estén conceptualmente vinculadas a los derechos y beneficios de la persona y de la sociedad, están en clases independientes. La categorización es un proceso que requiere estudiar el dominio de manera deductiva, es decir, determinar las clases de mayor alcance dentro de la temática elegida.¹⁶ De esa manera, no quedan claros los principios que se utilizaron en el proceso de categorización de las categorías de asunto de la Seguridad Social brasileña. Creemos que, cuanto más evidentes son los términos, más fácil será su localización y utilización. Destacar algunas áreas en detrimento de otras, en el mismo nivel jerárquico, hace que el vocabulario pierda la consistencia entre las relaciones. Proponemos pasar la clase de Seguridad Social a un nivel jerárquico arriba, poniéndolo en el mismo nivel de las otras clases que se mencionaron en el párrafo anterior.

Además, se encontraron inconsistencias conceptuales, relativas al postulado de la monorreferencialidad. Sobre este aspecto, observamos los términos “Previdência do regime estatutário” y “Previdência do servidor” como términos distintos, pero que poseen una relación de equivalencia. Observamos todavía que el término “Previdência do regime estatutário” no se encuentra fácilmente en su dominio de la Seguridad Social; es más recurrente el término “Previdência

do servidor”. La relación de equivalencia no está expresada en el vocabulario analizado, aunque esté descrita en la metodología de elaboración del VCGE. Luego, conforme lo que se observó en el párrafo anterior, hay dos términos distintos para representar un único concepto y, así, se pierde el control lingüístico de los términos, uno de los principios básicos de cualquier vocabulario controlado. Aunque en el lenguaje natural no existan sinónimos perfectos, debemos buscar en el lenguaje especializado el control de sinónimos o casi sinónimos.¹⁷

Una de las características de las taxonomías es que no incluyen definiciones para los términos, solamente para las relaciones jerárquicas.¹⁸ El VCGE no posee definiciones para los términos, aunque se solicite, en la pestaña “Sugestões” (sugerencias), la colaboración de los usuarios del vocabulario para que compongan las notas de alcance (NA) con definiciones de los términos. Además, las definiciones constan en la metodología como una de las etapas de la elaboración del vocabulario en cuestión. Comprobamos, entonces, la voluntad de volver el VCGE todavía más específico y con relaciones mucho mejor definidas. Sugerimos, así, añadir a los términos las notas de alcance, que facilitarán la comprensión de cada término representado, proveyendo las características del concepto que permitirán su agrupamiento, e indicar las relaciones que traen seguridad al organizador del tesoro en el momento de establecer las relaciones.¹⁹

Vislumbramos que la utilización del vocabulario puede ir más allá de un mapa para navegación en páginas electrónicas del sector público brasileño. Su uso puede extenderse a otros propósitos, como por ejemplo: utilizarse por el Estado para representar el dominio —mapa conceptual— de cada sector público, enseñando su trayectoria a lo largo del tiempo; o puede emplearse por los ciudadanos en la des-

cripción, recuperación y reutilización de informaciones gubernamentales, atendiendo así la organización necesaria a la publicación de los datos abiertos.

Nuestras propuestas pretenden contribuir a que el VCGE se constituya como un instrumento dinámico de organización del conocimiento y que pueda responder a fines diversos, de acuerdo con los intereses de cada usuario. Con el desarrollo de los conceptos, de la estructura y de las relaciones entre los respectivos elementos, este instrumento podrá volverse, en el futuro, uno de los pilares para que se elabore una ontología en el ámbito del gobierno brasileño.

DISCUSIÓN

Cuando se nos presenta el VCGE, reconocemos algunos puntos que convirtieron el vocabulario en inconsistente, principalmente para el empleo que se pretendía utilizar, es decir, en la relación semántica entre los datos publicados por Dataprev.

Es necesario definir vocabularios que representen los datos publicados, de manera que garanticen la posibilidad de lectura por máquina y de intercambio entre aplicaciones. Un vocabulario que se aplica a la estructura del VCGE es el *Simple Knowledge Organization System* (SKOS), que puede utilizarse conjuntamente con RDF (Resource Description Framework) para representar semiformalmente sistemas de organización del conocimiento, como tesauros, taxonomías y esquemas de clasificación.

Es necesario que haya el apoyo de los órganos relacionados para el empleo práctico de las tecnologías de la Web Semántica, con el propósito de que la publicación de los datos abiertos del sector público brasileño se vuelva efectiva

y reutilizable por los ciudadanos. La racionalización de los esfuerzos, a través del empleo de instrumentos de organización del conocimiento ya existentes, puede impulsar el desarrollo de nuevas investigaciones que harán viables la mayor participación de la sociedad en lo que atañe a la transparencia y a la *accountability* de las informaciones. Por lo tanto, necesitamos establecer asociaciones entre los órganos gubernamentales, las iniciativas privadas y la sociedad para la publicación y la reutilización de los datos abiertos, porque éstos se vuelven más poderosos cuanto más se usen.²⁰

El Plan de Acción para la Infraestructura Nacional de Datos Abiertos (INDA) prevé la prospección de ontologías y vocabularios existentes que se reutilizarán en la publicación de datos gubernamentales abiertos. El sitio datos.gov.br funciona como un catálogo colectivo de datos de los más diversos órganos públicos. De esa manera, los “datos del catálogo se clasificarán y agruparán conforme los temas (asuntos) que figuran en el Vocabulario Controlado del Gobierno Electrónico (VCGE).”²¹ Aunque exista la expectativa de utilizar el VCGE en la organización semántica de los datos publicados, se demostró que el vocabulario en cuestión necesita revisarse y alterarse, a fin de perfeccionar la organización conceptual de las temáticas relacionadas.

Ante lo expuesto, en el contexto de utilización del VCGE como instrumento de organización del conocimiento empleado en la publicación de los datos abiertos del Estado brasileño, se fundamenta nuestra propuesta de revisión de la clase de la Seguridad Social. Al utilizar la metodología y los resultados aquí presentados, es necesario que otros órganos públicos, además de Dataprev, decidan analizar, proponer alteraciones y actualizaciones en las clases cuya temática contemple los datos gestionados por ellos.

CONSIDERACIONES FINALES

Observamos la existencia de leyes que garantizan el acceso a las informaciones públicas, pero estas informaciones se diseminan de manera dispersa y con poca utilización por la sociedad. Se trata de recursos poco explorados todavía, debido a la dificultad de búsqueda, procesamiento y reutilización.

La población que produce datos también los consume. Por ello, el Estado y, más específicamente, la Seguridad Social necesitan perfeccionar los servicios ya existentes y, al mismo tiempo, elaborar otros nuevos que contribuyan en una relación de simbiosis entre ella y la sociedad.

Sin embargo, todavía deben vencerse muchos retos, tales como: ausencia de vocabularios controlados y de modelos conceptuales, deficiente cooperación entre las áreas públicas para compartir experiencias, falta de semántica entre los datos, e ineficaz difusión de los datos ya publicados.

Ante lo expuesto, esperamos, en trabajos futuros, conseguir enseñar la aplicación de lo que se propuso al VCGE y estimamos que otras clases reciban la misma atención aquí dispensada a la Seguridad Social. Ese vocabulario puede reutilizarse con muchos fines en el ámbito gubernamental; basta con que otros pares también se aficien en su desarrollo y a favor de la sociedad en general.

REFERENCIAS BIBLIOGRÁFICAS

1. A. Chahin *et al.* (2004), *E-Gov.br: a próxima revolução brasileira: eficiência, qualidade e democracia: o governo eletrônico no Brasil e no mundo.*

El proceso de publicación de los datos gubernamentales abiertos...

2. J. C. Vaz; M. M. Ribeiro; R. Matheus (2010), “Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil”, en *Cadernos PPG-AU/UFBA*, 9, pp. 45-62
3. E. Méndez; J. Greenberg (2012), “Linked Data for Open vocabularies and HIVE’s Global Framework”, en *El Profesional de la Información*, 21 (3), pp. 236-244.
4. C. J. S. Ribeiro; R. F. Almeida (2011), “Dados Abertos Governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade”, Encontro Nacional de Pesquisa em Ciência da Informação, 12, 2011 (Brasília), *Anais...*
5. w3C Escritório Brasil (2009), *Melhorando o acesso ao governo com melhor uso da Web*.
6. C. J. S. Ribeiro; R. F. Almeida (2011), “Dados governamentais abertos...”, *Op. cit.*
7. T. Berners-Lee (2009), “Putting government data online” [en línea], <http://www.w3.org/DesignIssues/GovData.html>
8. J. Hendler; O. Lassila (2001), *The semantic web. Scientific American*.
9. T. Berners-Lee (2009), “Putting government data online”, *Op. cit.*
10. E. C. Aganette; L. Alvarenga; R. R. Souza (2010), “Taxonomias corporativas: um estudo sobre definições e etapas de construção fundamentado na literatura publicada”, Encontro Nacional de Pesquisa em Ciência da Informação, 11, 2010 (Salvador), *Anais...*
11. M. L. A. Campos; H. E. Gomes (2008), “Taxonomia e Classificação: o princípio de Categorização”, en *DataGramZero: revista de Ciência da Informação*, 9 (4).

El acceso a la información gubernamental...

12. C. H. Marcondes; M. L. A Campos (2008), “Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação”, en *PontodeAcesso*, 2 (1), pp. 107-136.
13. D. L. McGuinness (2003), “Ontologies come of age”, en Dieter Fensel (ed.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*.
14. H. E. Gomes (Coord.) (1990), *Manual de elaboração de tesauros monolíngues*.
15. *Anuário Estatístico da Previdência Social* (2012), AEPS 2011, 20.
16. M. L. A. Campos; H. E. Gomes (2008), “Taxonomia e Classificação...”, *Op. cit.*
17. H. E. Gomes (Coord.) (1990), *Manual de elaboração de tesauros monolíngues*, *Op. cit.*
18. J. A. Moreira González (2011), *Linguagens documentárias e vocabulários semânticos para Web: elementos conceituais*.
19. H. E. Gomes (Coord.) (1990), *Manual de elaboração de tesauros monolíngues*, *Op. cit.*
20. G. Byrne; L. Goddard (2010), “The strongest link: libraries and linked data” en *D-Lib Magazine*, 16 (11/12).
21. Brasil. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação (2012), *Plano de ação da Infraestrutura Nacional de Dados Abertos*.