

INTEROPERABILIDAD DE LA INFORMACIÓN EN EL GOBIERNO ELECTRÓNICO

E-GOV INFORMATION INTEROPERABILITY

SILVANA TEMESIO | Licenciada en Bibliotecología. Analista de sistemas. Magister en Documentación Digital. Profesora en la Universidad de la República (UDELAR), Escuela Universitaria de Bibliotecología y Ciencias Afines (EUBCA) de las materias base de datos y redes y sistemas.

RESUMO

O artigo analisa a gestão da informação, a comunicação e a extração de conhecimento no governo eletrônico do Uruguai, a partir de casos específicos: processos, documentos internos e aplicações informáticas, à luz do acoplamento terminológico. Apresenta-se uma estratégia de informação facetada com um enfoque transversal que aporta uma visão global e permite a oferta de dados abertos *ad hoc*. A integração semântica de dados é vista como uma linha de trabalho a ser desenvolvida nos bastidores das instituições governamentais.

Palavras-chaves: governo eletrônico; informação; dados abertos; integração semântica de dados.

ABSTRACT

This article analyses information management, communication and knowledge extraction in the Uruguayan e-government through specific cases: records, documents and applications, in light of terminology coupling. The article presents a faceted information strategy with a transversal approach which provides a holistic view and allows for supply of *ad hoc* open data. Semantic data integration is seen as a trend to be developed behind the scenes of government institutions.

Keywords: e-government; information; open data; semantic data integration.

RESUMEN

Se analiza la gestión de la información, la comunicación y la extracción de conocimiento en el gobierno electrónico en Uruguay, a partir de casos específicos: expedientes, documentos internos y aplicaciones informáticas, a la luz del acoplamiento terminológico. Se presenta una estrategia de información facetada con un enfoque transversal que aporta una vista global y permite la oferta de datos abiertos *ad hoc*. La integración semántica de datos es vista como una línea de trabajo a desarrollar en la trastienda de las instituciones gubernamentales.

Palabras clave: gobierno electrónico; información; datos abiertos; integración semántica de datos.

INTRODUCCIÓN

El gobierno electrónico es un cambio de foco que coloca al ciudadano como protagonista, lo insta a participar, a usar servicios, a interactuar y a acceder a los datos y a la información que las instituciones de gobierno generan. El gobierno es un gran productor de información y datos, y la iniciativa de gobierno abierto y las leyes de transparencia ofrecen un marco para su acceso. No obstante existen obstáculos de dispersión y diversidad de enfoques que plantean la necesidad de realizar una reingeniería enfocada en la información y la comunicación para que el acceso y uso efectivo pueda llevarse a cabo adecuadamente.

Se analizarán los siguientes casos en el marco de organismos de gobierno uruguayo:¹ documentos administrativos o jurídicos de tramitación bajo normativas regulatorias o jurídicas; documentos de uso interno que no integran el circuito jurídico y/o administrativo: buenas prácticas, informes, procedimientos, reportes que no fluyen sino que permanecen en reservorios donde son consultados; aplicaciones o sistemas informáticos.

A los efectos del análisis se utilizará el modelo de entidad relación (MER), diagrama de flujo de datos (DFD), diagrama de estados UML (Unified Modeling Language), análisis de integración semántica y una metodología visual que denominamos “mapas de información”.

Se analizan los temas de normalización terminológica, medulares para la integración de datos e información. Se proponen estrategias para analizar la integración de las aplicaciones informáticas en una institución y se muestran soluciones para asimilar visiones facetadas utilizando la idea de la integración de bases de datos federadas, formalizando una nueva exposición de datos abiertos.

La generación de conocimiento y la comunicación en la organización son temas que integran esta discusión atravesando todos los aspectos tratados. Finalmente se plantea una línea de trabajo hacia la elaboración de una ontología de gobierno electrónico.

PANORAMA GENERAL DE LA INFORMACIÓN

En la figura 1 se pone de manifiesto cómo se relacionan los datos, los documentos y los funcionarios que actúan sobre ellos, en cada caso con un rol específico en una institución de gobierno. La organización cuenta con información tácita que no está documentada y es detentada por los funcionarios que la adquieren en el desempeño de sus actividades para lo cual integran grupos de trabajo que elaboran documentos con diferentes roles: productores, revisores, lectores. Los funcionarios pertenecen a distintas oficinas que coordinan sus actividades. Las aplicaciones informáticas manejan datos y los grupos de trabajo que manejan las aplicaciones tienen distintos roles: escritores, revisores, lectores.

¹ El planteo tiene generalidad, pero se realiza a partir de la realidad concreta en Uruguay.

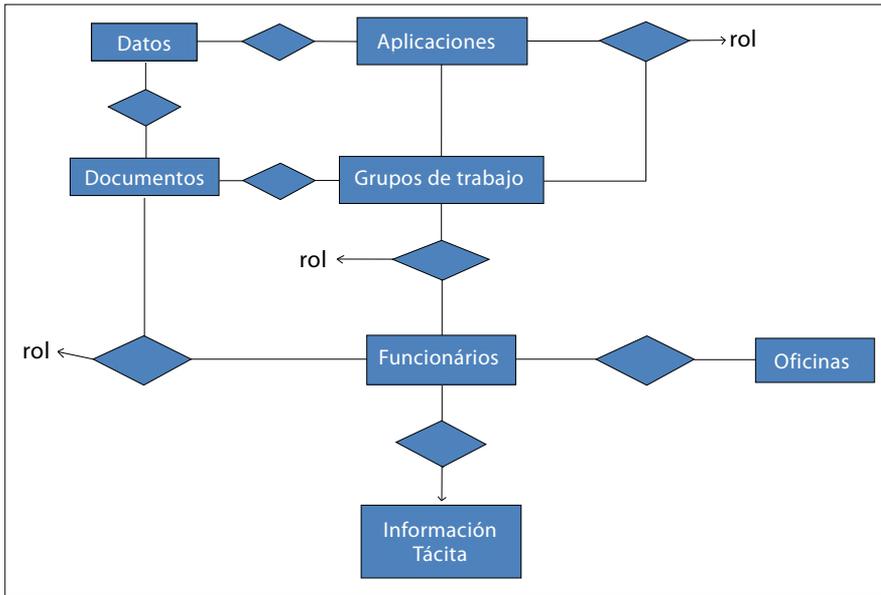


Figura 1. MER información egobierno

DOCUMENTOS ADMINISTRATIVOS

No se pretende realizar definiciones formales sobre los trámites administrativos sino analizar los aspectos generales desde un punto de vista abstracto. Para ello se toma como documento emblemático el expediente y se detallan sus aspectos relevantes a través de un diagrama de modelo entidad relación.

El expediente desde un punto de vista operacional es un documento que va fluyendo a través de la organización. Como se aprecia en la figura 2, va pasando por distintas instancias o estados en los que va sufriendo cambios, en rigor, va anexando fojas. En cada estado del expediente se permiten determinadas acciones llevadas a cabo por actores específicos. Veamos por ejemplo, un trámite licitatorio que se inicia con la confección de un pliego.

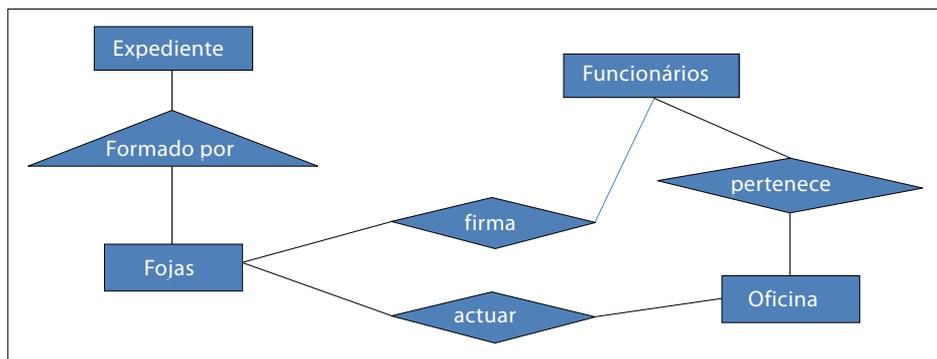


Figura 2. MER Expediente

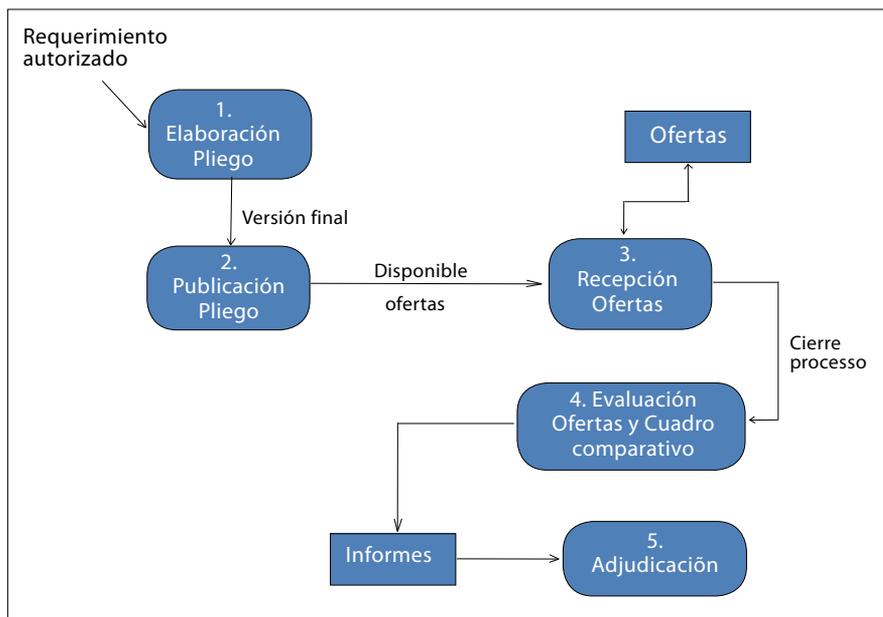


Figura 3. DFD trámite compra

El pliego tiene generalmente una parte técnica o de especificaciones y una parte legal. Mientras el pliego está en etapa de elaboración no se trata de un expediente, sino que es un documento que está siendo confeccionado por un grupo de trabajo: personal técnico, legal y administrativo. Cuando se finaliza, el documento es revisado por las jerarquías y sufre cambios que se ven reflejados en distintas versiones hasta que finalmente está pronto y puede darse inicio al trámite administrativo y entonces ese documento ya no cambia y va siguiendo los pasos que la tramitación administrativa especifique.

El documento inicial – el pliego – tiene un valor en sí mismo y puede ser objeto de consulta, independientemente del proceso administrativo de la compra. Este documento es una referencia de especificaciones que puede ser tomado como punto de partida para operaciones similares, en el mismo organismo o en otros. Una vez que se inicia el trámite, el documento inicial o pliego aún puede ser consultado, pero no es muy práctico porque está inmerso en el expediente y éste está fluyendo en la organización llevando a cabo su tramitación.

Respecto al trámite, el proceso de compra que se inicia con la elaboración de un pliego sufre una serie de transiciones como se observa en la figura 3. Una vez que la jerarquía autoriza el trámite licitatorio, hay un primer proceso que es la elaboración del pliego de especificaciones. Una vez concluido el proceso 1 de elaboración del pliego, hay un flujo de información que es la versión final del pliego que es insumo del proceso de publicación del pliego (proceso 2). Una vez que este proceso termina, se abre una instancia en la cual el pliego está disponible para la recepción de ofertas por parte de los proveedores (proceso 3), con el archivado de las ofertas. Vencido el plazo de recepción de ofertas, se dispara el proceso de evaluación de ofertas y elaboración del cuadro comparativo (proceso 4), el cual culmina con

la elaboración de un informe de recomendaciones para la adquisición, y dicho informe dispara el último proceso que es la adjudicación de la compra al proveedor elegido (proceso 5).

El proceso de compra se puede ver entonces como una serie de documentos hilvanados con actuaciones diversas en periodos y situaciones establecidas hasta que en algún momento el expediente se cierra y ya no puede haber actuaciones. Esto es una simplificación, porque puede haber instancias de impugnación o se puede reabrir el expediente por alguna circunstancia, pero son circunstancias que no agregan al análisis.

El expediente cerrado y completo luego de tramitado se archiva.

Pueden verse dos momentos bien diferenciados:

1) El expediente está en un flujo de procesos y va cambiando de estados, pero todavía el estado no es el final, y puede haber acciones y cambios en el expediente.

2) El expediente ya ha llegado en forma provisional a un estado final y se archiva.

En el primer caso normalmente existe una aplicación que va registrando el cambio de estado y los actores que intervienen (aplicación de seguimiento de expedientes). Si la aplicación además puede registrar los cambios en el propio documento que es guardado en formato digital y es firmado en forma electrónica estaríamos en el caso de la aplicación de expediente electrónico.

En el segundo caso, cuando el expediente es archivado (ya no fluye en la organización), entonces el documento completo o sus partes son descritas con las normas archivísticas: ISAD(G), ISAAR(CPF) y se consigna en una aplicación de descripción archivística como por ejemplo, ICAAtM. En este caso ya no interesa ver la situación en que está el trámite (v.g. disponible para ofertar) sino situar el expediente en el contexto, dentro de la organización, testimoniando su historia y devenir a través de acciones documentadas.

Como se ve, son dos instancias o situaciones diferentes con objetivos claramente distintos. En la tramitación interesa el estado, quien puede actuar, como sigue etc.; mientras que en el documento de archivo me interesa el contexto de la organización, quienes participaron con roles de importancia, de que se trata etc. En el archivo el documento es testimonial, no procedural.

Quiero destacar entonces, que hay visiones y momentos diversos sobre los mismos objetos. En el caso de la confección de un pliego por un grupo de trabajo, estamos hablando de una gestión documental. Si se decide guardar un registro para reutilizar a posteriori, podríamos hablar de un repositorio de documentos internos de la organización, cuando fluye el trámite hay una gestión de expedientes o expediente electrónico y finalmente cuando se archiva hay un tratamiento archivístico.

DOCUMENTOS DE USO INTERNO

Existen otros documentos fuera del circuito de tramitación a nivel de instituciones de gobierno, como informes, presentaciones, reportes de datos, mapas y procedimientos entre otros.

Los documentos están asociados a funcionarios que tienen distintos niveles de acceso al mismo, algunos pueden leerlos mientras que otros los escriben. La lista de control de acceso (access control list – ACL) es un concepto que se utiliza en las aplicaciones informáticas y en

los sistemas operativos, que maneja tres elementos: roles, privilegios y personas. Los roles se forman dando privilegios: se puede leer, se puede escribir, y estas acciones se aplican a determinados objetos (v.g. documentos). Luego las personas son puestas en los roles y queda establecido el nivel de acceso. Los documentos entonces tienen su ACL, habrá documentos internos de la organización, o de la oficina, o informes solo visibles a nivel de las jerarquías, y todas las combinaciones de visibilidad sobre objetos que se requiera. Definir una ACL sobre los documentos o carpetas de documentos permite modelar la accesibilidad de acuerdo a los requerimientos del caso o la aplicación de la ley de acceso a la información pública.

Los documentos que son producidos por un grupo de trabajo atraviesan distintas etapas y en cada una de ellas tienen actores con tareas y privilegios diferentes. En el momento de elaboración algunos usuarios tendrán el rol de escritor, otros de corrección, revisión o aprobación del documento final y luego de que el documento está finalizado y se publique otros usuarios tendrán el rol de lectura o consulta. Este proceso puede verse también como una máquina de estados, los usuarios escritores pueden modificar el documento pero cuando está en el estado de elaboración, una vez que el documento está publicado estos mismos usuarios tendrán por ejemplo el rol de lectura o consulta pero ya no el rol de escritura, de acuerdo a las transiciones que la gestión documental pauta.

Una vez publicado, el documento puede ir a un archivo pero también puede colocarse en un repositorio institucional, en una intranet o en un sistema de archivos en una red. El documento quedará disponible mientras tenga una utilidad, para ser usado como referencia, consulta o ser reusado.

Los documentos que tienen esta manipulación se tratan con una clase de aplicaciones que se denominan gestores documentales. Para fijar ideas podemos hablar de aplicaciones de software libre como Nuxeo o Alfresco, que ya tienen estas prestaciones establecidas pero sobre los cuales es posible establecer desarrollos nuevos que se adapten a transiciones específicas no contempladas.

Puede darse el caso de que se use un gestor documental en las diversas transiciones que sufre el documento, pero una vez finalizado el proceso, el documento se puede incorporar a otro tipo de aplicación como por ej. un repositorio. En el repositorio, el documento se ubica en una colección de acuerdo al uso previsto y se define una lista de control de acceso, por ejemplo colecciones de normativas o resoluciones para consulta general. En este caso, los repositorios son reservorios de documentación organizada por colecciones o temas que tienen un valor de uso más extendido en el tiempo y que son descritos con Dublin Core e indizados temáticamente con cierto nivel de normalización terminológica usando tesauros propios. Para fijar ideas podemos hablar en este caso de la aplicación Dspace para gestión de repositorios.

APLICACIONES INFORMÁTICAS

Las aplicaciones informáticas gestionan los datos crudos de la organización. Estas aplicaciones producen reportes que pueden verse en el contexto de la aplicación y también fuera

de ella. Los datos crudos están almacenados en bases de datos, generalmente separados en esquemas diferentes, pero también existen datos comunes que pueden ser usados desde distintos esquemas y aplicaciones.

En realidad todos los ítems documentales citados también contienen datos crudos, pero en el caso de las aplicaciones estos datos son tratados en forma sistemática, están ubicados dentro de una estructura y son identificables. Los datos además son incorporados a través de un proceso de validación tanto del tipo de dato como de determinadas condiciones semánticas que se establecen en las aplicaciones. Las aplicaciones también involucran funcionarios que las usan con distintos roles (lector, escritor etc.) y niveles de acceso a la información de acuerdo a los roles y tareas que se designen. En un sistema de gestión de personal, por ej. los jefes tendrán acceso a los datos de los funcionarios de su oficina, pero no los datos de los funcionarios de otras oficinas. También las aplicaciones pueden verse como una máquina de estados: hay opciones en los menús para ingreso de datos, opciones para consulta, opciones para procesos y todos son estados diferentes de los datos y tienen sus listas de control de acceso diferentes. Los datos son ingresados por unos usuarios y luego pueden ser revisados y aprobados por otros y luego son consultados por otros usuarios u otras aplicaciones. Aparece aquí un asunto nuevo que no hemos considerado, que es que los datos pueden ser ingestados por otras aplicaciones para lo cual tiene que haber una interoperabilidad semántica, es decir los datos intercambiados se deben corresponder en el significado.

APLICACIONES GRÁFICAS

Dentro de la categoría de aplicaciones o sistemas informáticos están las aplicaciones que son más específicas en los aspectos gráficos y dentro de las cuales quiero señalar a su vez dos grandes clases, sistemas que gestionan mapas y sistemas que gestionan fotos o gráficos.

SISTEMAS QUE GESTIONAN MAPAS

Los sistemas de información geográfica (SIG) presentan datos crudos provenientes de una fuente externa (por ej. una base de datos) sobre un mapa. En un mapa donde aparecen puntos que simbolizan ciudades, estos puntos tienen asociados datos crudos como nombre de la ciudad, cantidad de habitantes etc. Los datos se agrupan por capas sucesivas para poder gestionarlos en forma flexible y combinarlos u ocultarlos según convenga a lo que se desee hacer. Por ejemplo, podría haber una capa de ciudades, una capa de puentes, una capa de carreteras y así sucesivamente se van generando capas de información. Hay dos aspectos en un SIG: el "dibujo" y los datos asociados al "dibujo". A nivel de estado se promueve la normalización del "dibujo". Si todos tienen un mapa común será posible asociar capas generadas por distintos productores que ponen sus datos sobre el mismo "dibujo". Esta iniciativa es la IDE: infraestructura digital espacial y en cada país los distintos organismos coordinan para consensuar todos los elementos que son necesarios para trabajar en forma conjunta. Una IDE es una iniciativa que reúne acuerdos políticos, relacionamiento de organizaciones, recursos humanos, tecnologías, datos y servicios estandarizados de locali-

zación, visualización y descarga, que posibilitan el acceso e intercambio a diferentes niveles de uso de la información geográfica.²

Cada mapa o cada capa de un mapa puede asociarse con un conjunto de datos que lo describen, los metadatos geográficos. De esta manera los metadatos me informan, por ejemplo, el formato del recurso a acceder, con que frecuencia se actualiza o el método de representación espacial. Los metadatos geográficos usan la norma ISO 19.115.

Uno de los metadatos que describen los mapas son los descriptores y éstos pueden importarse de un tesoro usando el estándar SKOS (Simple Knowledge Organization System). El descriptor permite acoplar los mapas con otros tipos de datos en otros formatos – documentos, expedientes, aplicaciones etc. Si existe un tesoro institucional que provee terminología, entonces constituye un factor de interoperabilidad en la medida que se usa una terminología común y consensuada. Esta terminología es usada para recuperar un mapa, un documento que está en el sistema de gestión documental o en el archivo, siempre que cualquiera de estos recursos contengan como descriptor buscado una entrada en el tesoro.

SISTEMAS QUE GESTIONAN FOTOS O GRÁFICOS

Los sistemas que gestionan fotos o gráficos tienen adosados metadatos. Los metadatos son datos sobre la imagen, datos descriptivos y administrativos que contextualizan información sobre la imagen. Esto es muy importante porque los metadatos nos permiten por un lado encontrar la imagen específica en un mundo que es cada vez más gráfico, pero además brindan información propia de la imagen, como el autor de una foto o los datos técnicos de la fotografía, posibilitando la contextualización del recurso. Normalmente se usa Dublin Core o Dublin Core Extendido como estándar de metadatos. Dublin Core tiene un núcleo que es usado en forma estándar y se tiene la posibilidad de extender este núcleo dando flexibilidad de adicionar datos específicos usando la característica de extensión de los metadatos. Dublin Core brinda interoperabilidad permitiendo compartir recursos a través de metadatos comunes usando el protocolo OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting).

Puede darse la situación inversa que se tengan datos crudos con una imagen asociada (v.g. datos de medición de fracturas en el pavimento a los cuales se les adosa la foto de la fractura), en este caso se trata de aplicaciones informáticas que además de datos crudos enlazan imágenes que complementan el dato. En este caso se pueden guardar los gráficos desde la aplicación informática en la propia base de datos o ésta puede dar un enlace al gráfico que está en un sistema de archivos.

Pongamos por ejemplo una aplicación de gestión de carreteras que guarda datos del pavimento y que puede remitir a una aplicación gráfica que contenga fotos de fracturas de pavimento (v.g. piel de cocodrilo). Ambas aplicaciones deben compartir un lenguaje terminológico común: la aplicación de gestión debe usar formas normalizadas para nombrar las

² Disponible en <<http://www.agesic.gub.uy/innovaportal/v/665/1/agesic/IDE.html>>.

carreteras (ruta 1) y también formas normalizadas para nombrar las tipologías de fractura de pavimento (piel de cocodrilo) y lo mismo la aplicación gráfica. La normalización terminológica se construye con una aplicación específica que es el tesoro y debe ser coordinada y consensuada a nivel institucional.

TRATAMIENTO TERMINOLÓGICO

Los puntos de acceso son importantes en cualquiera de los ítems detallados: expedientes, documentos generados o usados por la organización y aplicaciones informáticas. Los puntos de acceso no solamente deben estar previstos para toda la tipología de objetos sino que tienen que ser homogéneos para la institución. Sería deseable una homogeneidad terminológica a nivel del Estado para estandarizar y facilitar la recuperación de la información y proporcionar un acceso genérico a nivel de los portales del Estado. El primer paso será lograr este objetivo a nivel institucional y en las interfases internas o intranets.

Cada institución tiene su terminología específica pero cuando se relacione con otra institución remitirá a la terminología de la institución de competencia. Cuando el Ministerio de Turismo se refiera a las rutas deberá usar los términos que proporciona el Ministerio de Transporte que es el que se ocupa de la adecuación terminológica en esta área (v.g. ruta 1), pero será el responsable de los términos que se refieran al turismo.

La descripción terminológica debe cumplir estándares (SKOS) y estar debidamente documentada para que pueda cumplir los objetivos planteados.

La organización, primero hacia la interna debe tener un lenguaje claramente establecido y documentado. Este lenguaje será usado para la recuperación de las categorías detalladas: documentos, gráficos o multimedia, aplicaciones etc. Este lenguaje de la organización debe elaborarse en forma participativa con el aporte de los usuarios técnicos que conocen la jerga de uso y su alcance y los profesionales de información. Este lenguaje debe ser claramente documentado y además ser ampliamente difundido, conocido y manejado en la organización para que resulte de utilidad.

Supongamos que el Ministerio de Transportes y Obras Públicas tiene un portal institucional interno (Intranet) y esta Intranet disponibiliza diversos formatos y objetos de información (mapas, archivos, repositorios etc.). Si se busca en la Intranet por ejemplo "ruta 1" debería recuperar un mapa digital con un valor "ruta 1" para el atributo "rutas" o las tareas de mantenimiento realizadas en una aplicación de mantenimiento de rutas que tenga el valor "ruta 1" para el campo "rutas" o el expediente con la actuación que hizo la empresa vial en la ruta 1 que figura con el descriptor "ruta 1" como punto de acceso, y todos estos resultados de la búsqueda tienen un elemento que nuclea y homogeiniza, que es una entrada en un tesoro para toda la organización donde bajo la jerarquía "rutas" figura la "ruta 1" con este nombre que se adopta como el término usado.

En este caso lo que se estandariza a nivel de la organización (y sería deseable estandarizar a nivel del Estado) es el valor del punto de acceso, aportando un nombre único y un significado claro y establecido que elimine ambigüedades.

Es necesario dar un paso más: los atributos no solo tienen un valor, también tienen una definición de trabajo. A medida que las aplicaciones se vuelven más especializadas como los sistemas de información geográfica (SIG) o aplicaciones a medida que usa la organización resulta necesario contar con una definición de trabajo clara y concisa. Una consulta muy común a nivel vial es el tránsito promedio diario anual (TPDA) en una ruta. Para ver exactamente que significa el valor TPDA debemos tener claro que tipos de vehículos están definidos en este valor, y así poder compartir, analizar, comparar y en definitiva trabajar con esta información. Es necesario establecer claramente el alcance del término y consignarlo en un tesoro u ontología.

MAPAS DE INFORMACIÓN

Para analizar la información a nivel institucional se plantea el uso de una herramienta que en adelante se denomina “mapa de información” y es un bosquejo de alto nivel de los ítems productores de información y la forma como se relacionan unos con otros. Esta herramienta es gráfica y por tanto tiene la bondad de mostrar claramente como son los vínculos y relaciones de las aplicaciones de todo tipo y sus productos en una institución. El mapa de información es un grafo dirigido en el cual los nodos son los objetos productores de información, por ej. aplicaciones informáticas y los arcos indican los flujos de información que se intercambian entre los nodos y sus características (conexión manual o automática, tipo de información).

Resulta evidente la notoria ventaja de tomar como insumo la información que producen otros y centrarse en darle valor a aportando otros elementos que la enriquecen y no tener que generar nuevamente la misma información, con ineficiencia y redundancia. El mapa de información visibiliza los insumos y salidas de información en un entorno de trabajo definido, por ejemplo una institución.

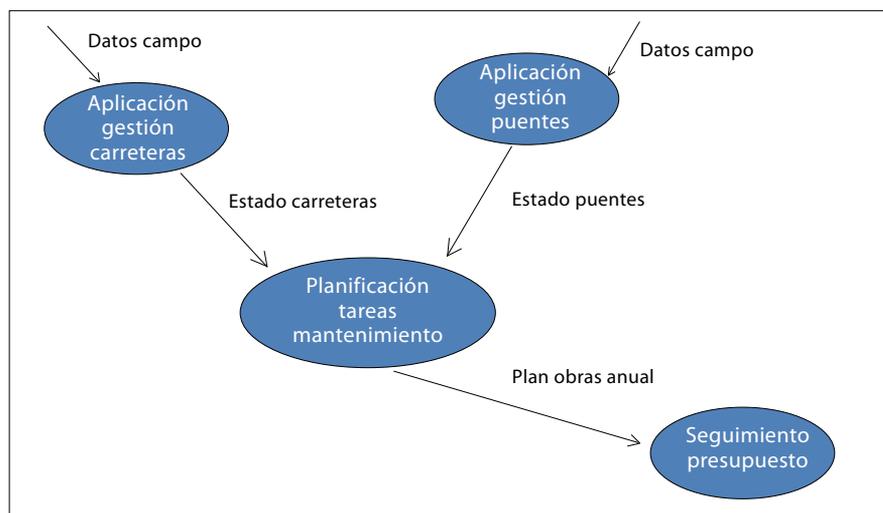


Figura 4. Mapa información

Un ejemplo de mapa de información en el campo de la gestión vial se puede ver en la figura 4. En el caso ejemplo, existe una aplicación de gestión de carreteras que tiene como entrada, datos de las carreteras que se relevan en el sitio (en campo) y constituyen un insumo para esta aplicación que tiene una salida que es el dato del estado de las carreteras que a su vez es un insumo para la aplicación que realiza la planificación de las tareas de mantenimiento de carreteras.

Los flujos de información que conectan las aplicaciones pueden ser manuales o automáticos. Lo ideal es que existan articulaciones automáticas entre las distintas aplicaciones y que las salidas de aplicaciones que son insumos de otras se realicen en forma automática a través de mecanismos que garanticen la corrección de los acoples (*web services*, triggers de base de datos etc.). El flujo de información "estado puentes" es un atributo "puente" y un atributo "estado" con sus respectivos valores (Puente 1 = bien, Puente 2 = regular).

Aunque a primera vista parece claro lo que es un puente, existen conceptos muy cercanos (puente y alcantarilla) que pueden generar dudas. Entonces establecemos el alcance en un tesoro u ontología: *Puente*: un puente es una construcción que permite salvar un accidente geográfico o cualquier otro obstáculo físico como un río, un cañón, un valle, un camino, una vía férrea, un cuerpo de agua, o cualquier otro obstáculo; *Alcantarilla*: una alcantarilla es una construcción que permite el drenaje transversal de las aguas superficiales bajo otra infraestructura, como un camino, carretera o vía férrea. Funcionalmente, una alcantarilla es totalmente similar a un puente, pero el término se aplica a obras de porte relativamente pequeño.

Al brindar una definición de trabajo y establecer claramente los términos, podemos realizar un acoplamiento entre la aplicación "Gestión de puentes" y la aplicación "Planificación tareas de mantenimiento", que tomará acciones diversas de acuerdo a consideraciones técnicas, según se trate de puentes o alcantarillas. Luego de definido el atributo, deben establecerse claramente los valores. Si la aplicación "Planificación de tareas de mantenimiento" opera en puentes con estado="regular", se debe tener claro que significa estado="regular". El estado de los puentes debe ser inequívoco para que los acoples funcionen adecuadamente. Los acoples pueden ser manuales o automáticos, es decir las salidas de una aplicación pueden ser ingestadas directamente en otra aplicación en forma automática con mecanismos como disparadores de base de datos o puede suceder que haya que ingresar en forma manual los valores de salida de un sistema a la entrada del otro. En el ejemplo, para realizar una planificación anual de tareas de mantenimiento es necesario como insumo el estado de los puentes (proveniente de la aplicación de gestión de puentes) y si las dos aplicaciones no tienen un acople automático entonces un operador manualmente debe ingresar los datos. Hay que considerar que en el ingreso manual hay un nivel de error que es necesario estimar y tomar acciones correctivas para subsanarlo. También hay que considerar la ineficiencia de tener que ingresar valores que ya se tienen.

Para poder realizar un acople automático, precisamos no solamente un desarrollador que realice la aplicación de acople, sino establecer interoperabilidad semántica.

La interoperabilidad semántica correspondería al establecimiento de una concordancia clara y precisa de los significados de los datos que se quieren acoplar. La interoperabilidad

está en un modelo de capas, donde tendríamos: primer nivel, de interoperabilidad a nivel de infraestructura (las aplicaciones se pueden comunicar físicamente); segundo nivel, de interoperabilidad sintáctica (los datos que acoplo son compatibles en su tipo y valor); tercer nivel, de interoperabilidad semántica (los datos que acoplo son compatibles en su significado).

Si la aplicación “Gestión de puentes” no distingue entre puente y alcantarilla y la aplicación “Planificación tareas mantenimiento” solo trabaja con puentes, entonces no hay manera de reutilizar los datos y hay que verificar dato por dato para cada ítem si se trata de un puente o una alcantarilla. Por otra parte si se decide actuar sobre los puentes en estado “regular” pero no coincide la semántica de puente en estado regular para la aplicación “Planificación de mantenimiento” y “Gestión de puentes” entonces las acciones no serán certeras.

CONOCIMIENTO

La información precisa constituye un insumo para la toma de decisiones, para planificar, proyectar y establecer políticas públicas extrayendo conocimiento de esos datos e información.

Si la información está dispersa, será muy difícil analizarla para extraer conocimiento. La minería de datos es una metodología de extracción de conocimiento a través de aplicaciones específicas a esos efectos. Una solución de este tipo requiere planificación, esfuerzos y tiene su costo, pero sin duda permite establecer escenarios y realizar un análisis en prospectiva para tomar decisiones. Una solución alternativa que se plantea aquí, es generar aplicaciones que levanten los datos crudos de los distintos esquemas de bases de datos de la organización y los conecten de manera flexible estableciendo filtros y mostrando los resultados en forma gráfica.

Al realizar un análisis de la información de la organización se busca articular los datos al elaborar una capa de aplicación que permita manipular la visualización de la información utilizando filtros. El impacto de la visualización en formato gráfico y la combinación con el uso de filtros resulta muy revelador.

La riqueza expresiva de esta solución está dada por la rápida conexión de datos, saltando de una faceta a otra. Una consulta sobre una ruta arroja que está en buen estado (esquema de datos viales), entonces se puede visualizar cuanto se invirtió (con la articulación con un esquema de datos de sistema financiero) y saltar a ver cuales fueron las tareas de mantenimiento de rutas que se desarrollaron (esquema de gestión de carreteras) y así se van combinando las distintas informaciones y se puede ver que al realizar determinadas acciones sobre un tipo de pavimento específico se obtiene un resultado determinado con una inversión determinada, mientras que con otras metodologías hay otros resultados. Todo esto permite contrastar resultados para evaluar técnicas, inversiones y políticas.

En la figura 5 se muestra el esquema conceptual de la arquitectura de sistemas federados, que fue tomada como base para la propuesta. En nuestro caso, cada base de datos corresponde a un esquema de una aplicación y tiene una mirada sobre un aspecto de un objeto global, sería el esquema local. Este esquema local se lleva a un esquema componente

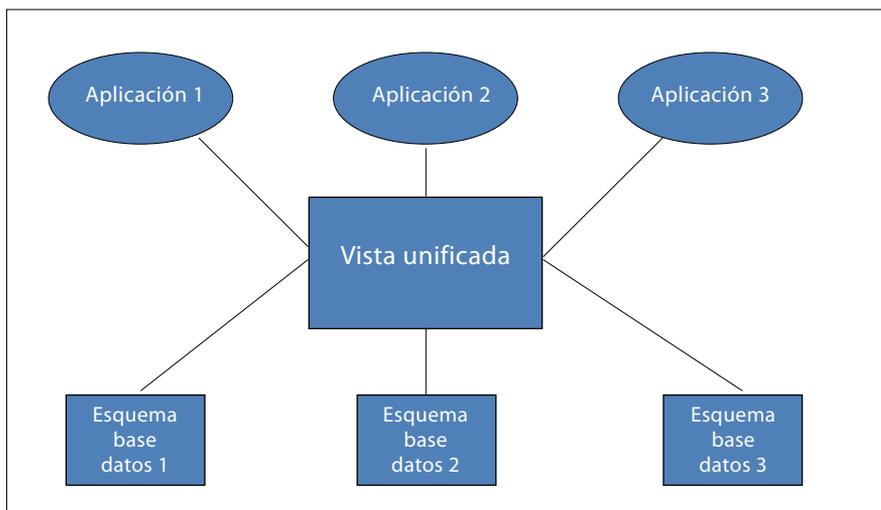


Figura 5. Esquema datos federados

que corresponde al modelo de datos común – a veces es necesario generar adecuaciones y disparadores que generen tablas nuevas. Desde este esquema componente se trabaja directamente ya que el esquema federado es la concatenación de todos los esquemas componentes y esta concatenación agrupa componentes prácticamente disjuntos. El elemento común es el identificador de un objeto principal sobre el cual se miran facetas diferentes. La integración de estas miradas diversas de un objeto común se teje directamente desde la aplicación de visualización, de modo que el esquema federado constituye una capa abstracta que tiene el conocimiento de estos enlaces. Los distintos esquemas externos se ven como un esquema externo único, una gran visión global que se navega hacia todas partes con filtros y enlaces.

En cierto sentido es una especie de *datawarehousing* que no bucea en características desmenuzadas y particionadas sino en muchos atributos diferentes, desde una perspectiva globalizadora. La visión no es en profundidad de atributos, sino en atributos distintos, facetas diferentes del mismo objeto.

Esta aplicación de visión unificada permite moverse con distintas opciones y aplicar filtros, ésta es una característica de importancia para poder hacer una exploración que genere conocimiento en forma ágil. A su vez cada atributo usado en la aplicación se detalla en su alcance y definición brindando documentación de estructura y de semántica. También se documenta la fuente del dato, es decir de que aplicación procede de modo de tener clara la trazabilidad del dato.

Es de destacar que una vez hecha una consulta el resultado es exportable en el formato que se desee: planilla electrónica, CSV (Comma-separated values) o RDF (Resource Description Framework), de modo que se puede decir que esta aplicación de visión unificada es una fuente de datos abiertos, con la característica de que no es necesario actualizar el conjunto de datos, ya que siempre el conjunto de datos está actualizado, pues muestra los datos que

están en los esquemas locales. La otra ventaja es que el conjunto de datos se genera *ad hoc*, de acuerdo a los requerimientos o necesidades, aplicando los filtros.

La aplicación de visión unificada, constan de un menú indicando distintas facetas de la información y para estas facetas ofrecen filtros con ventanas desplegadas y opciones gráficas. Los filtros muestran opciones que son intuitivas pero que en un análisis pormenorizado deben ser claramente establecidos y aquí está presente la necesidad de una claridad terminológica y es por eso que el sistema debe mostrar la documentación sobre la estructura y la definición de los atributos. Contar con esta documentación es una condición ineludible para la legibilidad del dato. Esta documentación corresponde al conocimiento extraído de la aplicación fuente.

Una vez que se identifica un dato de interés se puede ir navegando para verlo en detalle cada vez mayor. Se va desde lo general a lo particular, bajando cada vez más el nivel de detalle. La potencialidad más significativa de la aplicación de visión unificada es su disponibilización en forma gráfica y la posibilidad de exportar los datos en formato de planilla electrónica para usarse como datos abiertos.

SERVICIO UNIFICADO DE INFORMACIÓN

El diseño de información o la arquitectura de la información a nivel del gobierno electrónico es una tarea en la cual es necesario un grupo interdisciplinario de profesionales de la información (bibliotecólogos, archivólogos, documentalistas) en conjunto con informáticos y con el aporte de los licenciados de comunicación para el diseño de una política de comunicación. El fortalecimiento de todas las herramientas mencionadas en este documento y su disposición en un portal interno que sea liderado por este equipo interdisciplinario y que sea a la vez el artífice de recoger la información dispersa, articularla y disponerla estimulando a los funcionarios a participar en esta tarea puede establecer un cambio hacia la trastienda del gobierno electrónico. Este equipo también puede ser el encargado de recoger las consultas internas o externas provenientes del ejercicio de la ley de acceso a la información pública. Esta “ventanilla única de consultas” permitiría identificar vacíos, necesidades, requerimientos, adaptaciones y en general ir adecuando la información y los servicios a nivel institucional, hacia la interna, pero también hacia el ciudadano haciendo posible también la coordinación con otros organismos de gobierno bajo un criterio de acoplamiento racional.

Existen herramientas de amplio uso en otros ambientes como la “mesa de ayuda” que generan una base de conocimiento y establecen un monitoreo estadístico de consultas para articular un ciclo de mejora continua que atienda las consultas insatisfechas y detecte vacíos a partir de los cuales se puedan habilitar cambios, mejoras, nuevos procesos, nuevos actores, o valor agregado a lo que ya exista. La base de conocimiento que genera este tipo de aplicaciones de mesa de ayuda, normalmente permite establecer consultas estacionales, recurrentes en las que se puede ir mejorando el desempeño o incluso anticiparse con servicios. Los pedidos de información pueden categorizarse por áreas, por tipo de solicitante, por niveles de urgencia; todo lo cual habilita el establecimiento de políticas y actividades que apoyen en definitiva una mejor gestión, porque la gestión es para el ciudadano, es el ciudadano el cen-

tro de la actividad y una detección de las inquietudes informacionales del ciudadano hace a un gobierno más cerca de las necesidades e intereses de sus usufructuarios y más apegado a la realidad social que es la que reclama la información.

CONCLUSIONES Y PROSPECTIVA

Las distintas funciones y actividades que lleva adelante una organización van dejando rastros tangibles en distintos objetos y formatos que se decantan en contenedores específicos (aplicaciones, colecciones digitales, documentos archivísticos etc.). Cuando esos objetos no son documentos archivísticos, es necesario trasuntar la vinculación respecto a esas funciones o actividades a través de metadatos específicos.

En el caso de los documentos archivísticos a diferencia de otras fuentes de información, éstos tienen la característica intrínseca de que se estructuran de acuerdo a la relación orgánica que los vinculan reflejando las funciones y actividades que documentan. En el documento archivístico la adición de metadatos es un elemento funcional a la integración semántica y a la recuperación, pero la ubicación del documento archivístico en la topología del archivo es un elemento semántico propio que preexiste en el cuadro de clasificación del archivo y que no se señala específicamente pero que no puede soslayarse ni sustituirse por este aporte de metadatos.

En la información que fluye en la organización tenemos datos no estructurados (contenidos en documentos de trabajo y expedientes), datos semiestructurados (contenidos en formularios o planillas Excel) y datos estructurados (provenientes de las bases de datos). Esta información es dinámica, va cambiando, enriqueciéndose, mudando de un sistema a otro y el mayor desafío es lograr acceder a todos los aspectos de una información que está en situaciones, formatos y sistemas diversos.

Los datos tienen un ciclo de vida y están sujetos a cambios a lo largo de las actividades que realiza la organización. El dato tiene una fuente, un origen, una trazabilidad. Si se precisa un dato sobre un atributo v.g. el estado de un puente, pueden interesarme los datos específicos que brinda una aplicación informática que utiliza una base de datos, pero luego quiero ver una foto del puente, que está en un repositorio de fotos, o preciso acceder a un informe de un técnico en un documento que está en un gestor documental o al expediente por el cual se contrató un mantenimiento del puente que está en el sistema de expedientes o en el archivo institucional. Para lograr enlazar esta diversidad, es necesario construir cañerías que me permitan deslizarme de un lugar a otro usando los mismos términos y sin tener que ser un experto en cada una de estas instancias. Para ello es necesario lograr la interoperabilidad terminológica y la integración semántica.

Es necesario generar consenso sobre el significado de los datos, su semántica. ¿Qué significa cuando yo hablo del estado de una carretera? Hay una definición de trabajo que opera en un contexto y que debo visibilizar para saber cual es el significado del término y poder usarlo apropiadamente.

Cada aplicación tiene autonomía, usa su propio modelo de datos, sus propios estándares y funcionalidades, pero existe la necesidad de acoplar estas aplicaciones y es entonces necesario

consensuar un lenguaje común en el que por ejemplo “estado” tenga el mismo el significado en una aplicación o en otra, o si esto no se logra, por lo menos debe quedar claro que no hay correspondencia semántica y que en cierto lugar un término tiene un significado y en otro lugar otro significado, para de este modo acoplar los elementos que puedan hacerlo o realizar los ajustes o modificaciones para que se acoplen los elementos que tienen diferencias.

En síntesis, la información en la organización es heterogénea, está dispersa, no está estructurada, es dinámica (sufre cambios a lo largo del tiempo) y usa una terminología diversa por lo cual realizar una búsqueda de información plantea un problema de integración de datos.

Podríamos llamarlo el problema del “metabuscador institucional” por analogía a los metabuscadores en los sistemas bibliotecológicos que reúnen en un solo punto las búsquedas en el catálogo bibliográfico online (Online public access catalog – OPAC), las bases de datos institucionales o contratadas, los repositorios, y una diversidad de recursos de información que posee una biblioteca.

En el gobierno electrónico existen diferentes fuentes de datos o información (repositorios, expediente electrónico, aplicaciones de bases de datos, sistemas de información geográficos, gestores documentales etc.) y tienen esquemas distintos en los cuales cada fuente manipula sus datos. Las fuentes son autónomas en cuanto a su diseño por tanto son heterogéneas en cuanto a su modelo de datos (heterogeneidad sintáctica) y son autónomas en cuanto a su manera de comunicarse.

Como la situación se aplica a un dominio específico – el gobierno electrónico – se puede hacer un abordaje que corresponde a un segmento de conocimiento acotado, estableciendo un modelo referente, consensuado. Esta es una ventaja importante porque además a nivel de gobierno hay una visión y una estrategia compartida en la cual existen lineamientos que propenden el establecimiento de vinculaciones interinstitucionales y servicios transversales.

Si existe una terminología para el dominio (nivel gobierno o nivel institución de gobierno) que sea consensuada, es posible mapear los términos de los distintos esquemas a esa terminología común.

Esta terminología que usa la organización está viva, se extiende en la medida que se realizan las actividades. Existe además un conocimiento en la organización que es rico en semántica y es necesaria una herramienta que permita describir la riqueza del dominio – relaciones de inclusión de conceptos, herencia etc. – enriqueciendo la terminología de modo que pueda reflejar esta compleja urdimbre. Esta herramienta es una ontología.

Para compartir información en un dominio, entonces, se debe tener un referente y un acuerdo sobre un modelo que sea común. Ese modelo común debe tener la característica de que se pueda modificar, extender, adaptarse a nuevos elementos o rediseñarse y eso se logra con una estructuración que sea algo más que una taxonomía consensuada o un tesoro, eso se puede hacer con una ontología que es extensible y que permite definir operaciones y realizar inferencias.

Como primera aproximación a la integración semántica se plantea trabajar con un tesoro de la organización, estableciendo una terminología común para luego seguir avanzando en la elaboración de una ontología en el ámbito de cada institución de gobierno. Las instituciones a su vez pueden coordinar unas con otras los elementos que corresponden a

la competencia de cada quien. La idea es apropiarse del espíritu de la IDE en lo referente a compartir áreas comunes y desarrollar áreas específicas. El Ministerio de Transporte se puede ocupar de los términos viales mientras que el Ministerio de Economía de los términos de presupuesto y también hay un segmento común a todo el estado como el de los expedientes que correspondería a un nivel central.

Las ontologías desarrolladas pasarían a ser un estándar de uso común del lenguaje del dominio – en este caso el gobierno electrónico –, de la estructuración de sus conceptos y sus relaciones, en definitiva constituyen una interpretación de la realidad en ese segmento de conocimiento establecida como común para construir interoperabilidad. Son un referente al que se pueda aludir y un modelo común para expresiones no canónicas que pueden seguir existiendo y desarrollándose con autonomía pero que pueden ser interoperables si existe un mapeo entre sus elementos y los de la ontología referente.

La ontología tiene un gran poder expresivo y además puede extenderse a medida que se requiera. La ventaja mayor es que se trata de una especificación formal por lo cual es posible extraer conocimiento y puede ser usada para consultas semánticas.

Si se puede establecer un mapeo entre las fuentes diversas y la ontología, es posible poner los datos de las fuentes (que están no solo semiestructurados, sino estructurados muy diversamente) y convertirlas en instancias de la ontología y es la ontología la que contesta las preguntas pudiendo extraer conocimiento nuevo.

El modelo común para expresar los modelos de datos de los distintos esquemas tiene que ser el de mayor expresividad, RDF, que es un modelo de datos conceptual que constituye una recomendación de la W3C para la web semántica. Se trata de triplas (sujeto, predicado objeto) que describen los recursos y se puede serializar en XML. Pueden verse como un grafo ordenado navegable. Una vez que se adhiere a un modelo común RDF para la descripción de los recursos, es necesario resolver la heterogeneidad semántica, las discordancias en el significado o interpretación de los datos estableciendo las correspondencias semánticas.

Finalmente se puede realizar búsqueda semántica, por ej. en SPARQL (SPARQL Protocol and RDF Query Language), que es un lenguaje de consulta sobre RDF.

Este camino es un trayecto de muchos tramos y de largo aliento pero la ventaja de iniciarlo a nivel de instituciones gubernamentales es que éstas fortalecen los procesos internos de conexión y recuperación de información. Al tratarse de un dominio restringido es una tarea de una escala razonable. El conocimiento de la organización se abre hacia sí misma mejorando las interconexiones internas y evitando errores y redundancia, y el conocimiento va quedando documentado y accesible, de modo que las fuerzas que se pierden en tareas que se repiten a lo largo de la organización pueden usarse para dar información más rica y correcta.

En la medida que las instituciones de gobierno transiten un camino similar – que se puede establecer como una política pública – es posible ir generando los acoples interinstitucionales y generar una normalización de información a nivel de gobierno que se puede ir completando al paso en que las organizaciones vayan avanzando.

El papel de los profesionales de la información en esta iniciativa, formando parte de equipos multidisciplinarios es fundamental.

Referências bibliográficas

BOOCH, G.; JACOBSON, I.; RUMBAUGH, J. *The Unified Modeling Language Reference Manual*. Madrid: Pearson Higher Education, 2004.

CALLEGARI, Daniel; VIERA, Marcos; MOTZ, Regina. Design of a service-oriented architecture for federated systems. *Journal of Computer Science & Technology*, v. 5, n. 4, 2005. Disponible en: <<http://journal.info.unlp.edu.ar/journal/journal16/papers/jcst-dec05-1.pdf>>. Acceso en: 27 mayo 2013.

CARBALLA RIVAS, Noa María. La comunicación 3.0 en el gobierno abierto: las redes sociales como estrategia de comunicación para la información cotidiana. CONGRESO INTERNACIONAL COMUNICACIÓN 3.0., III, 10 y 11 oct. 2012, Salamanca. Disponible en: <<http://campus.usal.es/~comunicacion3punto0/comunicaciones/2012/502.pdf>>. Acceso en: 27 mayo 2013.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. *e-ARQ Brasil: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos: versão 1.1*. Rio de Janeiro: Arquivo Nacional, 2011. Disponible en: <<http://www.documentoseletronicos.arquivonacional.gov.br/media/e-arq-brasil-2011-corrigido.pdf>>. Acceso en: 27 mayo 2013.

ELMASRI, R.; NAVATHE, S. B. *Fundamentos de sistemas de bases de datos*. 3. ed. Madrid: Pearson Higher Education, 2002.

MOTZ, Regina. Dynamic Maintenance of an Integrated Schema. *Integrated Publication and Information Systems to Information and Knowledge Environments*, p. 21-30, 2005. Disponible en: <http://link.springer.com/chapter/10.1007/978-3-540-31842-2_3#page-1>. Acceso en: 27 mayo 2013.

OECD. OECD Guiding Principles for Open and Inclusive Policy Making. EXPERT MEETING ON BUILDING AN OPEN AND INNOVATIVE GOVERNMENT FOR BETTER POLICIES AND SERVICE DELIVERY, 8-9 jun. 2010, Paris. Disponible en: <<http://www.oecd.org/gov/46560128.pdf>>. Acceso en: 27 mayo 2013.

OSZLAK, Óscar. Gobierno abierto: promesas, supuestos, desafíos. CONFERENCIA ANUAL INPAE 2012, VIII, 25 de abril de 2012, San Juan de Puerto Rico. *Gobierno abierto: por una gestión pública más transparente, participativa y colaborativa*. Disponible en: <<http://www.oscaroszlak.org.ar/images/articulos-espanol/Gobierno%20abierto.pdf>>. Acceso en: 27 mayo 2013.

PEDRAZA JIMÉNEZ, R.; CODINA, L.; ROVIRA, C. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, v. 16, n. 6, p. 569-578, 2007. Disponible en: <<http://www.lluiscodina.com/webSemanticaOntologias2007.pdf>>. Acceso en: 27 mayo 2013.

RAMIREZ ALUJAS, Alvaro. Innovación en la gestión pública y open government (gobierno abierto): una vieja nueva idea. *Buen Gobierno*, n. 9, p. 94-133, 2010. Disponible en: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1820326>. Acceso en: 27 mayo 2013.

_____. Gobierno abierto y modernización de la gestión pública: tendencias actuales y el (inevitable) camino que viene. Reflexiones seminales. *Enfoques*, n. 15, p. 99-125, 2011. Disponible en: <<http://www.gigapp.org/es/lista-de-miembros?view=publication&task=show&id=387#UaO9QkqbOso>>. Acceso en: 27 mayo 2013.

SHETH, Amit P.; LARSON, James A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys (CSUR)*, v. 22, n. 3, p. 183-236, 1990. Disponible en: <<http://knoesis.wright.edu/library/download/SL90.pdf>>. Acceso en: 27 mayo 2013.

Recebido em 28/5/2013

Aprovado em 10/7/2013