



1 de enero de 2017 | Vol. 18 | Núm. 1 | ISSN 1607 - 6079

ARTÍCULO

METODOLOGÍA DE INVESTIGACIÓN EN HUMANIDADES DIGITALES

(<http://www.revista.unam.mx/vol.18/num1/art01/>)

Silvana Temesio

(Magister en Documentación Digital, Universidad Pompeu Fabra ,España)

METODOLOGÍA DE INVESTIGACIÓN EN HUMANIDADES DIGITALES

Resumen

...puede plantearse entonces es que el área y sus líneas de investigación pueden enmarcarse dentro de las humanidades digitales.

Se presenta la metodología de investigación utilizando herramientas y técnicas que provienen del análisis de sistemas (análisis estructurado, MER, UML, DFD, ontologías, grafos, etc.) aplicadas al estudio de sistemas de información.

La aplicación de esa visión está fuertemente ligada al objetivo de la web semántica en un plano que globalice e integre la información aislada generando un conocimiento interconectado y trazable.

En particular se remite sumariamente la aplicación de esta metodología a tres casos:

- Documentos y datos en el gobierno electrónico.
- Documentación legislativa.
- Datos abiertos.

Los objetos de información que están a la interna de las instituciones en formatos diversos: aplicaciones, colecciones digitales, multimedia, archivos, datos, entre otros, pueden vincularse y relacionarse permitiendo tránsitos que logren valor agregado. El establecimiento de una semántica consensuada permitiría articular facetas dispersas y trabajar en la creación de un conocimiento integral de la información de estas instituciones. Esta línea de investigación puede enmarcarse en el espacio emergente de las humanidades digitales.

Palabras clave: web semántica, gobierno electrónico, humanidades digitales.

Research Methodology in Digital Humanities

This paper presents a research methodology that uses tools and techniques that come from systems analysis (structured analysis, MER, UML, DFD, ontologies, graphs, etc.). This methodology is applied to the study of information systems. The implementation of this vision is strongly linked to the goal of the semantic web in the sense that aims globalization and integration of the isolated information to transform it into traceable and interconnected knowledge. This methodology is briefly referenced on the application on three areas in particular:

- *Data and documents in e-government.*
- *legislative documentation.*
- *open data.*

Information objects that are internal to government institutions in various formats: applications, digital collections, multimedia, files, data, archives, among others, can be linked

and related allowing paths that achieve added value. The establishment of consensus over semantics will allow articulate dispersed facets and strive on the creation of a comprehensive knowledge of government information. This line of research can be framed in the emerging space of digital humanities.

Keywords: *Semantic Web, e-government, Digital Humanities.*

METODOLOGÍA DE INVESTIGACIÓN EN HUMANIDADES DIGITALES

Introducción

El presente trabajo plantea la utilización de metodologías y técnicas que provienen del ámbito del análisis de sistemas aplicados al estudio de sistemas de información y comunicación. Se utilizan estas metodologías para desmenuzar la información, el conocimiento, los flujos en que participa y los aspectos de comunicación a la interna en el sistema o instituciones, con el foco en el paradigma de las ciencias de información: la recuperación de la información y la extracción de conocimiento.

La aplicación de esta visión está fuertemente ligada al objetivo de la web semántica en un plano global que integra la información aislada generando un conocimiento interconectado y trazable.

El proceso de construcción de software parte de analizar un problema o situación y buscar la solución. A partir de este análisis es que se diseña la solución.

Los modelos de desarrollo de aplicaciones en cascada son metodologías de análisis estructurado (*Structured Systems Analysis and Design Method*) que se basan en fases secuenciales que se alimentan unas a otras. Se avanza a una fase posterior a medida que se concluye la actual. Existen algunas variantes de este modelo con cambios en las fases pero con el criterio secuencial de empezar una fase una vez concluida la anterior. Una variante es la que considera la repetición del ciclo cuando se llega a la última fase en un proceso iterativo de refinamientos sucesivos. Una de las características de esta variante es que desarrolla un proceso iterativo ya que a partir del testeado o mantenimiento de la solución se realizan ajustes que requieren nuevos relevamientos o análisis de la situación lo que habilita nuevamente el ciclo: una nueva solución que a su vez será nuevamente examinada o ajustada.

Respecto a las metodologías de desarrollo de Sistemas de Información (Kendall, 2005) comenta que:

el análisis y diseño de sistemas busca comprender que se necesita para analizar la entrada o flujo de datos de manera sistemática, procesar o transformar los datos, almacenarlos y producir información en el contexto de una organización específica. Mediante el análisis detallado se busca identificar y resolver los problemas correctos. Además, el análisis y diseño de sistemas se utiliza para analizar, diseñar e implementar las mejoras en el apoyo para los usuarios y las funciones de negocios que se puedan llevar a

cabo mediante el uso de sistemas de información computarizados.

El ciclo de vida del desarrollo de sistemas (SDLC) es una metodología en fases para el análisis y diseño de los sistemas que comprende las siguientes fases:

1. Identificación de los problemas, oportunidades y objetivos.
2. Determinación de los requerimientos humanos de información.
3. Análisis de las necesidades del sistema.
4. Diseño del sistema recomendado.
5. Desarrollo y documentación de software.
6. Prueba y mantenimiento del sistema.
7. Implementación y evaluación del sistema.

En contraste, las metodologías *Rapid Application Development* (RAD) son más flexibles y se basan en desarrollos de prototipos. La idea es un desarrollo colaborativo, con una capacidad de evolución y adaptación alta con una continua mejora.

Los métodos ágiles son una serie de metodologías para el desarrollo de sistemas que se centran en los usuarios. Se puede mencionar a SCRUM como una metodología ágil que se arma con iteraciones incrementales en las cuales se va formulando el proyecto a lo largo de instancias de comunicación. XP o *extreme programming* hace énfasis en la interacción entre destinatarios y desarrolladores que interactúan en forma estrecha.

El proceso unificado de desarrollo (RUP) se centra en la utilización de casos de uso que el usuario espera implementar. Un caso de uso describe los requerimientos funcionales en términos de los distintos actores que interactúan con el sistema. Los actores representan un tipo de usuario del sistema. (Kroll, 2003)

También están las metodologías o técnicas orientadas a objetos que usan las herramientas de especificación UML (Unified Modeling Language). UML aporta un método sistemático para abordar el análisis y diseño orientado a objetos (Craig, 1999). Los objetos son entidades abstractas que tienen características y operaciones que constituyen la interfase con otros objetos, otras clases y con el mundo.

La primera fase del ciclo SDLC se centra en la identificación de los problemas, con lo cual podríamos abordar el planteo también desde la psicología aplicada a la resolución de los mismos.

(Simon, 1971) establece las siguientes fases para la resolución de problemas:

- Inteligencia.
- Diseño.
- Elección.

En la fase de inteligencia se identifica el problema y se recolecta información al respecto. La fase de diseño desarrolla posibles soluciones al problema y en la fase de elección se decide una alternativa de las planteadas en el diseño.

Extensiones posteriores añadieron algunas fases:

- Inteligencia.
- Diseño.
- Elección.
- Implementación .

- Revisión.

La implementación es el desarrollo de la elección y la revisión es la evaluación de la decisión.

Como se puede observar hay cierta analogía con el enfoque de diseño estructurado.

Antecedentes

La aplicación de estas técnicas en el ámbito de las ciencias de información no es nueva. Podemos citar como antecedente el uso del Modelo Entidad Relación en FRBR (O'Neill, 2011) o UML en FRBRoo (Doerr, 2007).

Lo que no se ha podido encontrar como antecedente es la aplicación de estas técnicas en el estudio de un problema informacional desde la perspectiva del profesional de la información con una visión de integralidad en el marco de una metodología de investigación.

Web semántica

Dice la W3C (<http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>)

La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

En esta definición de la web semántica se han señalado algunos conceptos que refieren al quehacer del profesional de la información.

Dotar de semántica es un objetivo que se aplica a la web pero es igualmente aplicable al interior de las organizaciones, a nivel de los portales internos o aún en las aplicaciones de gran diversidad que se utilizan a la interna de las entidades: gestores de contenido, gestores documentales, colecciones digitales, aplicaciones de nicho, archivos, expediente electrónico, entre otros, a los cuales denominaremos contenedores de información. Estos contenedores de información almacenan información de diverso tipo, sujeta a lógicas y procesos propios. Estos contenedores eventualmente se vinculan entre sí para cruzar información o aportar facetas distintas sobre un mismo evento. Lo que se busca es la organización semántica de la información en el interior de una institución, dando semántica a los objetos informacionales que están a la interna de la institución en estos contenedores y generando caminos de vinculación entre esta información que está

aislada.

Para esclarecer este panorama se puede construir un mapa de información a nivel de una organización que contendrá todos estos contenedores diversos señalados y los objetos informacionales que son albergados por esos contenedores, así como los flujos de información entre los contenedores.

Un objeto informacional podría categorizarse en objeto simple u objeto complejo. La foto que coloco en el entorno virtual de aprendizaje calificaría como objeto simple, en el sentido que la información que encierra está básicamente recuperada en sus metadatos.

Como objeto de información compleja podríamos señalar el caso de un libro. En el caso del libro, la descripción bibliográfica me da información sobre el objeto que es "exterior" al mismo:

García, M. G. (2007). *Cien años de soledad*. Madrid: Alfaguara.

No obstante en el caso de un objeto de información complejo hay información que está en el "interior" que se pierde:

- ¿Dónde vivió José Arcadio Buendía?
- ¿Cómo se llaman sus hijos?
- ...

La descripción bibliográfica no se ocupa de estos aspectos que están en el "interior" del objeto informacional. Hagamos la salvedad del resumen analítico que describe el "interior" pero no parcela la información de manera buscable y semántica.

Aplicar los conceptos de la web semántica a un objeto de información complejo es establecer un grafo con conceptos que se vinculan con predicados y que es navegable e interrogable.

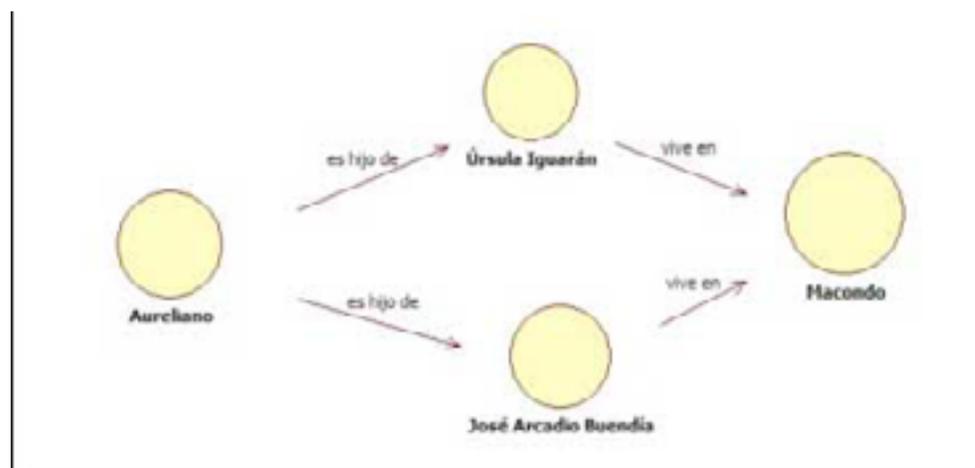


Figura 1. Grafo sobre la información "interior" de *100 años de Soledad*

Del grafo expuesto surge la necesidad de establecer un consenso terminológico sobre los conceptos. Es necesario establecer claramente un término que represente el concepto o una forma de delimitarlo por el contexto en que se utiliza.

Si se establece un término patrón, éste puede servir para navegar hacia otros objetos que puedan estar vinculados y en ese sentido tiene el mismo objetivo normalizador de un tesoro. Establecer una denominación del concepto Aureliano es de importancia para poder remitir variantes (Aureliano Buendía, Coronel Aureliano Buendía).

Por otra parte el grafo puede extenderse más allá del libro, a otros objetos que usen los conceptos que aquí aparecen o se relacionen con ellos. Por ejemplo el personaje de Aureliano está inspirado en el presidente colombiano Rafael Uribe Uribe, de modo que sería posible vincular una obra sobre este militar colombiano con el personaje de Aureliano. Así se va tejiendo una red de conceptos que sobrepasan los objetos y permite navegar y buscar en forma semántica y global.

La interfase de comunicación de la información en el gobierno electrónico hacia la ciudadanía son los portales institucionales. Allí se dispone la información:



Figura 2. Portal del Ministerio de Transportes y Obras Públicas



Figura 3 Código página del portal.

Además de la información que aparece en el portal se puede analizar el código html de la página en cuestión donde encontrar alguna información extra, sería el caso del título de la página por ejemplo. Sin embargo existe información que está al "interior" de la institución que no es accesible a través de los portales. Si se formulara una pregunta como por ejemplo que trabajos se hicieron sobre la ruta 1, encontramos que esta información no está disponible, o al menos no está disponible directamente.

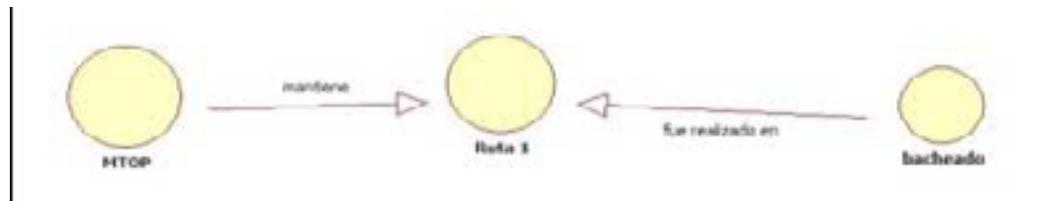


Figura 3a. Mantenimiento de la ruta 1.

Lo que se quiere establecer es que un objeto informacional complejo puede estar en una página web, en una colección digital, o en un sitio en el que puede ser encontrado, pero la información que contiene en su "interior" no puede recuperarse directamente.

La web semántica acomete la gigantesca labor de extraer información de los contenidos de las páginas web, pero las páginas web no son los únicos objetos del que se puede extraer información.

Lo que aquí se plantea es que los objetos de información que están a la interna de las instituciones de gobierno en formatos diversos: aplicaciones, colecciones digitales, multimedia, archivos, datos, etc. pueden vincularse y relacionarse permitiendo tránsitos que logren realizar extracción de conocimiento. Bajo esta perspectiva abordada en (Temesio, 2013) se estudia la información en dominios diferentes.

Casos estudiados

Vamos a citar tres artículos en los que se ha trabajado con esta metodología y con esta visión:

- Documentación en el gobierno electrónico (Temesio, 2013).
- Información legislativa (Temesio *et. al.*, 2014) .
- Datos abiertos (Temesio, 2015).

Caso gobierno electrónico

El caso analizado es el gobierno electrónico y se aplican técnicas provenientes del análisis de sistemas para el estudio de una solución o una propuesta de trabajo en un área de gobierno electrónico específico.

El gobierno electrónico se relaciona con aspectos como la gestión o administración, la toma de decisiones, la planificación estratégica, los procesos internos, la planifi-

cación, la elaboración de políticas públicas, entre otras. En particular se propone abordar los aspectos informacionales del gobierno electrónico utilizando las técnicas o metodologías del análisis de sistemas con un enfoque desde las ciencias de la información

La técnica de investigación que se ha utilizado para estudiar distintos aspectos del gobierno electrónico tiene una correspondencia con la metodología del diseño estructurado:

1. Relevamiento y análisis del problema.
2. Especificación formal representando la realidad.
3. Diseño o propuesta de solución.
4. Testeo o adecuación a lo largo del tiempo.

El uso de esta metodología se fue dando de forma natural, no planificada, más bien como una técnica de resolución de problemas en el abordaje.

Se han utilizado las técnicas de modelado lógico de datos (MER, UML), modelado de flujo de datos (diagrama de flujo de datos), modelado de eventos y procesos (diseño de eventos) y también se incorporaron elementos de la web semántica como RDF y ontologías. En todos los casos no se enfocaron en el desarrollo de aplicaciones sino en el análisis de las estructuras de datos que fluyen y se acoplan a través de procesos.

Las instituciones de gobierno son conglomerados de documentos, datos y aplicaciones, esto constituye un minimundo de información al cual es posible de aplicarle las metodologías de la web semántica, como un minimundo acotado y jerárquico es posible establecer procesos que hagan viable una información de gobierno semántica con mayor facilidad que en la web.

En el fondo del análisis emerge el problema de la terminología consensuada que permita dirigir los flujos y caminos adecuados y establecer operaciones bajo parámetros compatibles sobre las fuentes de información heterogéneas. Se trata del mismo paradigma de la web semántica.

El objetivo está en dar semántica a los objetos informacionales del mismo modo que en la web semántica y poder realizar una recuperación de información semántica y en múltiples formatos.

El gobierno electrónico tiene una exposición documental y de datos en los portales de gobierno, pero esto constituye una parte ínfima en cuanto a la información que existe en la trastienda, a la interna de la organización. La cual necesita además estandarizar la información para poder dar respuesta a la gestión y realizar planificación estratégica.

Información legislativa

Se analiza la información legislativa y se propone una ontología del dominio.

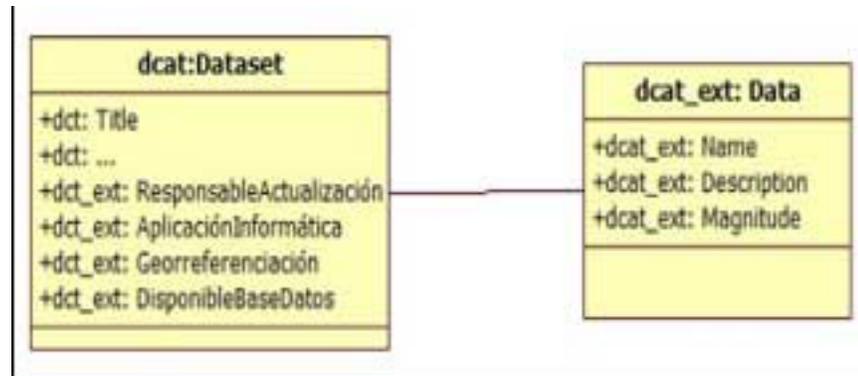


Figura 5. Extensión propuesta para DCAT

Se establece que:

Los datos de una organización son un elemento más en un entramado de otros objetos de información. Para que la información de una organización pueda acoplarse e interoperar es necesario estandarizar, sistematizar y documentar claramente estos datos en la gestión interna de la organización. Esta tarea también se debe realizar hacia fuera, interactuando con otros organismos gubernamentales y permitiendo el flujo de información interinstitucional. Un nivel superior es el uso de los datos de gobierno por parte de la ciudadanía, los periodistas u otros actores de la sociedad civil. (Ciudadanía/Ciudadanía 2.0).

Conclusiones

En el gobierno electrónico el universo de la información es muy amplio: tenemos documentos, aplicaciones, datos, informes, imágenes, mapas, actas, normativas, archivos, expedientes, etc. Podemos brindar información desde la bibliotecología o la archivología pero lo estaríamos haciendo desde un silo aislado. El enfoque planteado es brindar información en forma integral, global. Al aplicar la técnica del diseño de sistemas en la organización o el dominio y estudiar los procesos, los flujos de información, los almacenes de datos de la organización y como ellos se comunican y acoplan es posible lograr esta visión total.

Más allá de los aspectos específicos en los que se centran estas investigaciones y las que sigan en esta línea, lo que se plantea acá puede verse como un "mashup¹" interdisciplinar. Pero las características de este remixado en el que confluye una colaboración desde diferentes visiones puede verse además como expresión de una disciplina que pugna por emerger: las humanidades digitales. Las humanidades digitales justamente buscan la integración de los métodos computacionales en las ciencias humanas y lo hacen con una perspectiva crítica y reflexiva. (Orlandi, 2002)

Estas líneas de investigación entonces desde este punto de vista pueden verse como pertenecientes a las humanidades digitales.

Dentro del área de Documentación digital en el Instituto de Información en la Facultad de Información en Uruguay, se ha intentado poner en los objetivos y programas un enfoque que promueva la utilización de herramientas informáticas para la apropiación de los problemas disciplinares de las ciencias de información en un marco tecnológico. Lo que puede plantearse entonces es que el área y sus líneas de investigación pueden enmarcarse dentro de las humanidades digitales.

Los procesos y metodologías de investigación han surgido de manera natural en una simbiosis para el abordaje de las situaciones o problemas informacionales a estudiar, la perspectiva recién se revela en el momento de reflexionar sobre el marco teórico o los aspectos metodológicos. Puede decirse entonces que estos métodos, técnicas, modelos formales, entre otros, pueden constituirse en herramientas genéricas para el estudio de las cuestiones de información y comunicación en un enfoque de humanidades digitales y de las cuales los casos expuestos son solamente un ejemplo.■



[1] Mashup: es una forma de integración y reutilización. Ocurre cuando una aplicación web es usada o llamada desde otra aplicación, con el fin de reutilizar su contenido y/o funcionalidad. El uso en otra(s) fuente(s), para crear nuevos servicios simples, visualizado en una única interfaz gráfica diferente. Por ejemplo, se pueden combinar las direcciones y fotografías de las ramas de una biblioteca con un mapa para crear un mashup de mapa.(extraído de Wikipedia)

Bibliografía

- [1] *Ciudadanía/Ciudadanía 2.0*. [en línea]. España; Innovación Ciudadana en Iberoamérica, 2013. Disponible en: <http://www.ciudadania20.org/>. [Consulta: 25 de Noviembre del 2015].
- [2] LARMAN, Craig. *UML y Patrones. Introducción al análisis y diseño orientado a objetos*. Roig Vázquez, Pablo Eduardo (ed.). México: Prentice-Hall Hispanoamericana, 1999. 507 p. Traducción de: *Applying UML patterns: an introduction to object-oriented analysis and design*. ISBN: 970-17-0261-1.
- [3] DOERR, Martin. LEOEUF, Patrick. *Modelling intellectual processes: the FRBR-CRM harmonization*. En *Digital libraries: Research and development*. Berlin: Springer Verlag, 2007. (pp. 114-123). Lecture notes in computer science, 4877. ISBN: 9783540770879.
- [4] FERNÁNDEZ ALARCÓN, Vicenc. *Desarrollo de sistemas de información: Una metodología basada en el modelado*. [en línea]. Barcelona; Ediciones. UPC. 2005. 124 p. ISBN: 84-8301-862-4. Disponible en: <http://nubr.co/7ZOOb57>. [Consulta: 25 de Octubre del 2015].
- [5] JACOBSON, Ivar., BOOCH, Grady., RUMBAUGH, James. *El proceso unificado de desarrollo de software*. Sanchez, Salvador (trad). [en línea]. Madrid; México: Addison-Wesley. 2000. 438 p. Traducción de: *The Unified software development process*. ISBN:84-7829-0362. Disponible en: <http://nubr.co/LolGji> [Consulta: 25 de Octubre del 2015].
- [6] KENDALL, Kenneth., KENDALL, Julie. *Análisis y diseño de sistemas*. [en línea]. México: Prentice Hall, 2005. ISBN: 9786073205788. Disponible en: <http://nubr.co/E0OD6J> [Consulta: 25 de Octubre del 2015].
- [7] KRUCHTEN, Philippe. *The rational unified process made easy: a practitioner's guide to the RUP*. Mexico City: Addison-Wesley, 2003, 310 p. Addison-Wesley object technology series. ISBN: 0321197704.
- [8] O'NEILL, Edwards T. "FRBR: Functional requirements for bibliographic records". *Library resources & technical services*. 2011, vol 46, núm. 4, p. 150-159.
- [9] ORLANDI, Tito. *Is Humanities computing a discipline?* [en línea]. Roma: Centro Interdipartimentale di Servizio, 2002. Disponible en: <http://nubr.co/3a53GD>. [Consulta: 25 de Octubre del 2015].
- [10] SIMON, H. A. "The functional equivalence of problem solving skills". *Cognitive Psychology*. 1975, vol. 7, núm. 2, p. 268-288.

- [11] TEMESIO, Silvana. "Interoperabilidad de la información en el gobierno electrónico". *Revista Acervo*. [en línea]. 2013, vol.26, núm.2, p. 189-206. Disponible en: <http://nubr.co/HLCKDI>. [Consulta: 25 de Octubre del 2015].
- [12] TEMESIO, Silvana. "Metadatos de datos abiertos: estudio de caso en Uruguay". *BiD: Textos universitaris de biblioteconomia i documentació*. [en línea]. 2015, vol, 34 . Núm. 9. ISSN: 1575-5886. Disponible en: <http://bid.ub.edu/es/34/temesio.htm>. [Consulta: 25 de Octubre del 2015].
- [13] TEMESIO, Silvana., ESCUDER, Santiago., BARITÉ, Mario." La legislación y su visibilidad en sitios oficiales de Uruguay: aproximación crítica" En: *Uruguay digital ¿Cómo estamos? Múltiples miradas, un solo compromiso*. [en línea]. Disponible en: <http://nubr.co/MFw6uc>. [Consulta: 25 de Octubre del 2015].