# Organização da Informação no ambiente Web: produzindo conhecimento a partir de grandes volumes de dados

*Organization and information in the Web: producing knowledge based on large volumes of data*

**Cláudio José Silva Ribeiro**

UNIRIO – Universidade Federal do Estado do Rio de Janeiro, Av. Pasteur, 458/sala 413, CEP 22290-240 – Rio de Janeiro – RJ – Brasil – claudio.ribeiro@unirio.br

**Resumen**

O tema Big Data tem despertado interesse nos profissionais que trabalham com a Gestão da Informação, pois trata de insumo essencial no processo de criação do conhecimento. O estudo deste insumo sempre esteve evidenciado no ámbito da Ciência da Informação. Este relato apresenta o tema e explora os fundamentos que auxiliam no entendimento da abordagem de Big Data. Discute a explosão informacional e a avalanche de dados, chegando aos elementos que compõem o tema. Aborda os 5 Vs do Big Data e as fases de Discovery, Data Preparation, Model Planning e Analytics, bem como as características necessárias para desempenhar o papel de Cientista de Dados. O processo de pesquisa se baseia em um estudo exploratório desenvolvido por meio de pesquisa bibliográfica com análise documental e estudo de caso. O marco empírico para análise dos elementos do dominio partiu da análise da relevancia de itens publicados no site de Dados Abertos da Dataprev, onde foram extraídos os conjuntos de dados e as facetas identificadas. Apresenta os resultados preliminares sobre estudos nas fases iniciais de projetos de Big Data, de forma a viabilizar alternativas para representação e disseminação dos grandes volumes de dados presentes na Internet. Ao final, reúne alguns aspectos ligados ao perfil do profissional que está participando destes projetos.

**Palabras clave**: Big Data. Gestão da Informação. Representação da Informação. Facetas. Dados Abertos. Dataprev.

**Abstract**

The Big Data theme has been awakening the interest of professionals who work in the field of Information Management, as it deals with the essential input in the knowledge creation process. This input has always been a focus of study in the sphere of Information Science. This report presents the theme and explores the underpinnings of Big Data in order to achieve a better understanding of this approach. It discusses the information explosion and avalanche of data to single out the elements that comprise the theme. It analyzes the 5 V's of Big Data and the phases of Discovery, Data Preparation, Model Planning and Analytics, was well as the characteristics necessary to perform the role of Data Scientist. The research process was based on an exploratory study using a bibliographical survey, in addition to documental research and a case study. The empirical framework for analyzing the elements of the domain was developed by examining the relevance of the ítems published at the Dataprev's (Brazil's National Social Security Institute's Information Techonology organization) Open Data website, from which the data sets and facets identified were extracted. This study presents the preliminary results of studies in the initial phases of Big Data projects aimed at developing ways of representing and disseminating the large volumes of data found on the Internet. It ends by presenting some features of the profiles of professionals who are engaged in these projects

**Keywords**: Big Data. Information Management. Information Representation. Facets. Open Data. Dataprev.

## 1. Introdução

O uso de informação e conhecimento tem impulsionado projetos de investigação, tanto em ambientes acadêmicos quanto em organizações empresariais. Na visão de Castells (1999), nós presenciamos o surgimento de uma economia "informacional e global":

> A economia é informacional porque a produtividade e a competitividade de unidades ou agentes nessa economia (sejam empresas, regiões ou nações) dependem basicamente de sua capacidade de gerar, processar e aplicar de forma eficiente à informação baseada em conhecimentos. É global porque as principais atividades produtivas, o consumo e a circulação, assim como seus componentes (capital, trabalho, matéria-prima, administração, informação, tecnologia e mercados) estão organizados em escala global, diretamente ou mediante uma rede de conexões entre agentes econômicos. É informacional e global porque, sob novas condições históricas, a produtividade é gerada, e a con-

Ribeiro, Claudio Jose Silva. *Organização da Informação no ambiente Web: produzindo conhecimento a partir de grandes volumes de dados*. En XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de noviembre, 2015, Organización del conocimiento para sistemas de información abiertos. Murcia: Universidad de Murcia.

> corrência é feita em uma rede global de interação (Castells, 1999, p. 87).

Por outro lado, percebe-se que o uso de dados e informação sempre esteve evidenciado como objeto de estudos para a área de Ciência da Informação. As discussões sobre esses insumos essenciais na criação do conhecimento foram impulsionadas pela revolução científica e tecnológica que se seguiu à Segunda Guerra Mundial. No bojo dessas discussões, também marcadas pela necessidade de ajustes tecnológicos (Bush, 1945), surgiram novos conceitos como o "Caos Documental" ou ainda a "Explosão Informacional" (Ribeiro, 2008). Os profissionais envolvidos com pesquisas na área de CI sempre estiveram presentes em trabalhos ligados aos aspectos informacionais, mas sofrendo forte influência também pelos apelos tecnológicos (Saracevic, 1996a).

Na visão contemporânea trazida pelo Prof. Aldo Barreto:

> A chegada de uma sociedade eletrônica de informação modificou a delimitação de tempo e espaço dos conteúdos em relação aos receptores. Mas, a explosão de informação de que discorriam Vannevar Bush e Dereck de Solla Price no ambiente de pós guerra mundial foi minorada pelo computador no tempo possível.(Barreto, 2014, *online*)

Partindo de categorização sugerida por Barreto (2014), pode-se afirmar que estamos no "tempo do ciberespaço", onde a colaboração, a instantaneidade e a fluência digital cada vez mais se fazem presentes na Ciência da Informação. Surge a noção da "avalanche de dados", pois os dispositivos móveis, em especial os celulares e tablets, que atuam como geradores e consumidores de dados e informação estão presentes no nosso cotidiano. Esse movimento tem sido conhecido como Big Data e vem despertando o interesse por todas as pessoas que tem algum envolvimento com atividades para Gestão da Informação.

A possibilidade de fazer análise deste vasto volume de dados, vem ocasionando um incremento nos debates sobre este campo. Neste sentido, um exemplo recente sobre o uso dessa abordagem vem do esporte. Em reportagem veiculada por diferentes meios de comunicação sobre a seleção da Alemanha, vencedora da Copa do Mundo no Brasil, houve registro sobre a utilização de solução que permitisse a análise de dados sobre jogos, jogadores, táticas de jogo e a possibilidade de avaliar comentários especializados. Ou seja, o uso de dados e informação em grande quantidade como forma de melhorar o desempenho do time alemão. (InfoExame, 2014).

Desde Castells, percebe-se que a economia informacional propõe o uso de dados e informação na obtenção de resultados, logo, depreende-se que não estamos diante de uma novidade em pesquisas no campo da informação. A área de CI tem promovido estudos em temas que exploram esta temática, em especial em abordagens para a gestão de ativos de informação na WEB em níveis gerenciais, táticos e operacionais. (Ribeiro, 2008).

Dentro deste contexto, este relato apresenta alguns resultados preliminares sobre o envolvimento de profissionais da informação nas fases iniciais de projetos de Big Data, de forma a viabilizar alternativas para representação e disseminação dos grandes volumes de dados presentes na Internet.

## 2. A avalanche que incrementa o volume de dados

A maior disponibilidade de computadores, editores de texto, programas de correio eletrônico, a multiplicidade de redes sociais e de colaboração, a computação móvel, programas que fomentam a troca de mensagens, o forte uso do serviço Web e de localização geográfica, facilitaram em muito a criação ou obtenção de novos objetos portadores de informação. A tecnologia disponível na atualidade é extremamente adequada para produzir e armazenar múltiplas informações com vasto volume, contudo, nem sempre a organização destas informações permite a recuperação imediata de determinado conteúdo.

Cabe ressaltar aqui um outro aspecto: a forte relação entre dados e informação, que remonta a debates originados em várias áreas do conhecimento. A visão adotada nesta pesquisa é trazida por Davenport (1998), onde dados são simples observações sobre o estado do mundo, sendo facilmente estruturados, quantificados e transferíveis. Davenport continua e registra que a informação pode ser entendida como um conjunto de dados dotados de relevância e propósito, requerendo análise e mediação para se obter consenso em relação ao seu significado.

Neste contexto é possível afirmar que dados são necessários para a geração de informação e quanto maior for a quantidade de dados, maior será a quantidade de informação a ser tratada.

Ademais, como observado na introdução deste relato, a crescente utilização de meios de comunicação, com alto grau de mobilidade e fazendo uso da Web, contribuem para o aumento desse volume de dados. Percebe-se que o caos

documental e a explosão informacional continuam presentes, mas agora caracterizados como a "avalanche de dados", pois os dispositivos móveis e as câmeras de vídeo incrementam cotidianamente o volume de dados produzido e consumido pela sociedade. (Heath; Bizer, 2011; Shiri, 2014).

Estes aspectos endossam a previsão feita por Alvin Tofler, quando este autor observou as mudanças no comportamento da sociedade pós-moderna, em seu livro publicado em 1984. O termo *prosumer (1)*, forjado por Tofler, definia um novo perfil de interação da sociedade de consumo (Tofler, 1984).

Xie, Bagozzi e Troye (2008) e Fonseca, Gonçalves, Oliveira e Tinoco (2008) convalidam o surgimento deste novo perfil quando observam que esta sociedade passou a ser composta por individuos que possuem muita informação e estão constantemente buscando benefícios para seu próprio consumo. Consequentemente, é possível inferir que esta forma de interação começou a ser impulsionada a partir do incremento do uso dos dispositivos móveis que dão acesso imediato a uma miríade de informações

Em nível mundial, estima-se que a expansão das fontes de dados tenha crescimento de aproximadamente 50 vezes nos próximos 10 anos. (Emc, 2014).

Robredo (2011) também constata o crescimento e mostra que a quantidade de registros digitais passou de 0,28 ZB (zetabytes) (2) em 2007 para 1,2 ZB em 2010 e espera-se um enorme salto para 35 ZB em 2020.

Fruto deste cenário rico em volume e variedade de fontes, tem surgido uma disciplina que, apesar de não ser apenas um tema essencialmente tecnológico, vem sendo impulsionado pelos projetos de tecnologia: a vertente de Big Data.

## 3. Tratando a necessidade de informações para grandes volumes de dados

Compreender as necessidades de informação, parte do pressuposto que deverão ser mapeados os desejos e anseios dos usuários participantes de um domínio ou contexto específico. Nesta direção, procura-se estudar o que leva os usuários a formular perguntas, bem como são formuladas as respostas a estas perguntas. (HjØrland; Albrechtsen, 1995).

No entanto, tratar grandes volumes de dados torna complexa a compreensão das necessidades, uma vez que pode-se estar diante de múltiplas facetas contendo diferentes informações. A geração e coleta de muitos dados, além dos objetos portadores de informação que possuem formatos variados, nos remete para um contexto onde existem problemas que desconhecíamos por completo. A necessidade de gerir estes altos volumes repousa na abordagem conhecida como "Big Data" (Fox; Hendler, 2011).

Para dar conta dessa abordagem será preciso explorar ma\is alguns aspectos: a noção de Big Data, a visão de *Analytics* e o novo perfil que está surgindo para atuar neste campo, o Cientista de Dados.

### 3.1. Fundamentos sobre Big Data

A World Wide Web reúne fontes de diversas naturezas e, consequentemente, apresenta aos pesquisadores vários desafios para trabalhar com a imensa coleção de dados e informação em estoques. Barreto (2014, *online*) corrobora esta noção quando observa que "[..] com a condição online os estoques e os fluxos de informação, renomeados para 'Big Data', são multidirecionados [..]". Vencer os desafios originados pela avalanche de dados, exige para o profissional que trabalha com organização de informação competências diferentes no tratamento deste insumo, pois na visão de Cavalcanti (2014, online) "Para entender este novo mundo precisamos de novos óculos, de um novo pensamento (precisamos evoluir do pensamento cartesiano para um pensamento mais complexo) [..]".

Os projetos que usam a noção de Big Data precisam cobrir diferentes fontes e naturezas de dados, que podem ser entendidos como elementos ligados ao domínio da informação. Já para elementos ligados à tecnologia, estes projetos devem ser desenvolvidos fazendo uso de processamento em nuvem (3) e com tecnologias específicas, tais como processamento de rotinas em paralelo e ferramentas para otimização como *Hadoop (4) e MapReduce (5), HDFS (6),* além de abordagens de *MachineLearning (7).* (Davenport, 2014).

A noção de Big Data contempla o trabalho com alto volume de dados, com um processo de disseminação e atualização ocorrendo em grande velocidade. Além disto, tudo isto pode estar presente em múltiplas fontes e com alta complexidade para sua obtenção e análise (Dumbill, 2012).

Em artigo sobre o uso de Big Data com processamento em nuvem, Hashem et. al. (2014) complementam a noção, quando apresentam um conjunto de categorias que permite compreender as características de projetos com Big

Data. A compreensão das fontes disponíveis na Web de Dados, o originadas em projetos de colaboração que façam o uso de *social media*, além de levantar as características de formato dos dados e do processamento e/ou transformação destes, são aspectos que contribuem para o incremento de projetos com esta abordagem.

Inicialmente a noção de Big Data se apresentou apoiada em 3 características básicas, que são utilizadas para observar os diferentes aspectos envolvidos. São elas: o Volume, a Variedade e a Velocidade, caracterizando os 3 V's de Big Data.(Dumbill, 2012; Mello, Leite e Martins, 2014). No entanto, fruto do desenvolvimento de outras pesquisas conduzidas por empresas ligadas a projetos em Tecnologia da Informação, novas visões foram agregadas, sendo incorporado então um outro V para cobrir a parcela sobre a Veracidade dos dados. (Zikopoulos et al., 2013).

Power (2014) convalida a visão de Zikolopoulos et. al., observando que Big Data se apoia nos 4 V's, mas, de forma análoga, surgiu uma quinta característica essencial no tratamento dos dados e informação: o Valor. (Ibm, 2015). Segundo uma visão mais empresarial promovida por empresas de aconselhamento estratégico em Tecnologia da Informação (8), o ambiente de Big Data repousa sobre a possibilidade de transformar dados em informação que agregue valor aos negócios.

Assim, apoiado nas abordagens enunciadas anteriormente, pode-se dizer que para trabalhar com projetos de Big Data, será necessário observar as 5 dimensões presentes na Figura 1.

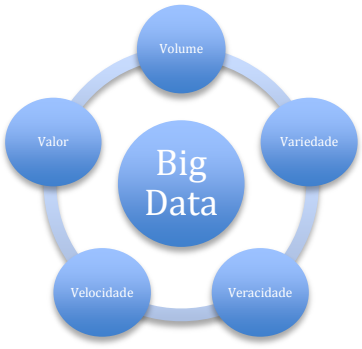


Figura 1: *Os 5 V's girando em torno de Big Data (Elaborado pelo autor a partir de Hashem et. al. (2014) e Ibm (2015)*

Duas das cinco dimensões presentes na Figura 1 correspondem a capacidade de tratar o grande quantitativo de dados presentes no cotidiano (Volume e Variedade). Como pode ser observado anteriormente (ver subseção 1), o uso de celulares e smartphones pelos indivíduos da nossa sociedade impulsionou o tema, na medida em que estes dispositivos podem funcionar com sensores que capturam e enviam informação. São cerca de 6 bilhões de dispositivos (OnuBr, 2014) gerando, consumindo, processando e (re)utilizando dados e informações.

Além destes dispositivos existem também câmeras de monitoramento em prédios, lojas, ruas e avenidas espalhadas pelas cidades. Este novo comportamento que vem sendo praticado pela sociedade, deixaria G. Wells, autor de 1984, impressionado com a possibilidade de observação que vem tomando a sociedade. A *Survaillance Society* (Mann; Nolan; Wellman, 2003) também vem incrementando o cotidiano com outra enxurrada de informações. Estima-se que a quantidade de vídeos produzidos diariamente ultrapassa a produção dos primeiros 50 anos de vida da televisão (Davenport, 2014).

Adiciona-se a esse cenário, uma vasta coleção de outras fontes e formas para geração de unidades documentárias. O crescimento do uso de documentos digitais e páginas Web nas organizações, recursos estes estruturados por meio de ferramentas para Gestão de Conteúdo (Ribeiro, 2008), bem como o desenvolvimento de propostas de uso da Web of Data e Linked Data (Ribeiro; Almeida, 2011) também têm contribuído para um aumento em Volume e Variedade de dados e informação.

A dimensão da Velocidade corresponde a característica mais perceptível para os usuários. A melhoria dos canais de transmissão, com redes em fibra ótica, emissores de sinais de alta capacidade e uso de satélites, traz reflexo direto no nosso cotidiano. Pode-se afirmar que o desenvolvimento da tecnologia de processamento, incluindo o uso de processadores, canais e hardware de armazenamento, dobra a cada período de 2 anos, o que contribui para incrementar o volume das unidades documentárias (Florissi, 2012).

A dimensão da Veracidade esta ligada à qualidade da informação (Abib, 2010). Esta é essencial para que os usuários interessados (executivos, gestores públicos e a sociedade em geral) usem e (re)usem os dados de maneira apropriada e real, gerando informações críveis e compartilhando significados. (Assis; Moura, 2011).

A dimensão do Valor é uma característica intrínseca dos dados, pois corresponde à capacidade desses dados gerarem informações que agreguem valor ao processo de tomada de decisão para os usuários. Essa capacidade está relacionada com formas de inovação e melhoria em processos de tomada de decisão, além de contribuir para a obtenção de vantagens econômi-

cas a partir destes grandes volumes de dados. (Power, 2014; Ibm, 2015).

### 3.2. Big Data *Analytics*

Trabalhar com as dimensões já apresentadas, pressupõe a possibilidade de analisar os dados e informações geradas. Neste sentido, sabe-se que a capacidade do indivíduo para identificar e representar requisitos e necessidades de informação é limitada. Esta capacidade depende, principalmente, da correlação entre os diferentes níveis de abstração e granularidade das representações no domínio que está sendo representado (Goguen; Jirotka, 1994).

Aliado a isto, é possível afirmar que estes indivíduos interpretam e representam de maneira diferenciada, pois a capacidade de percepção do cérebro humano estrutura a realidade e as conexões de informação segundo diferentes caminhos (Kent, 1998).

Assim, pode-se supor que para a execução da tarefa de análise desta vasta coleção, o ser humano precisa lançar mão de técnicas e ferramentas que reduzam a complexidade do domínio que está sob análise. A atividade de *Analytics* se caracteriza pela possibilidade de se realizar análise preditivas nos conjuntos de dados, fazendo uso de abordagens estatísticas (Siegel, 2013).

A separação em categorias e a posterior reunião em conjuntos, possibilita a execução de minerações (*mining*) (9) em determinados agrupamentos de grandes coleções, preparando assim os conjuntos de dados para serem carregados nos chamados lagos de dados (*data lake*). Este esforço de preparação é entendido como etapa de *Discovery* e precede a carga no *data lake.* (Oliveira, 2013; Tavares, 2014).

Na etapa de *Discovery* é possível utilizar outras abordagens para separação, limpeza dos dados e categorização, com a utilização, por exemplo, das fontes de origem do conjuntos (proveniência). Um dos objetivos finais para a etapa de *Discovery* é a obtenção de representações para os conjuntos carregados no lago de dados (*data preparation* e *model planning*). A Figura 2 representa de forma esquemática o ciclo para o trabalho com *Discovery*.

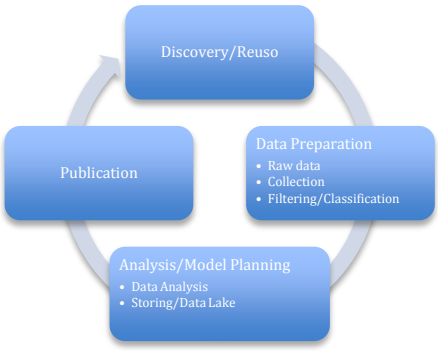


*Figura 2: Ciclo de Discovery (Elaborado pelo autor a partir de Tavares (2014) e Khan et. al. (2014))*

Este processo demanda uma nova arquitetura de análise, uma vez que os dados podem ser obtidos a partir de fontes diversas e originários de tecnologias diferentes. Seymour (2014) convalida esta percepção quando afirma que para analisar uma grande quantidade de dados, muitas vezes gerados de forma concomitante e em paralelo, pode existir uma grande dificuldade de correlacioná-los.

Ademais, Sathi (2013) destaca que em função destas múltiplas fontes, o trabalho com os processos de negócio de uma organização começa a prescindir de um profissional que saiba executar tarefas ligadas ao esforço de *Analytics*. Assim se torna possível desenvolver novos produtos e serviços para os clientes e/ou usuários.

### 3.3. O cientista de dados e o esforço de *Analytics*

A partir das observações apresentadas nas subseções anteriores é possível verificar a necessidade de desenvolver um novo perfil profissional que reúna diferentes disciplinas no trato de dados e informação. Decorrente desta inquietação, surgiu então o perfil denominado Cientista de Dados (*Data Cientist*). Davenport e Patil (2012) destacam que este profissional deve possuir, dentre outras características, a capacidade de aplicar ferramentas analíticas e algoritmos, com o intuito de gerar previsões para o desenvolvimento de produtos e serviços.

Na visão de Oliveira (2013), este profissional deve reunir conhecimentos em disciplinas ligadas à matemática e à estatística, principalmente pela necessidade de tratar os múltiplos conjuntos de dados. O uso de modelos matemáticos, a aplicação de técnicas de regressão, a capacidade de correlacionar variáveis e a formulação de hipóteses, também são características desejáveis para este profissional.

Oliveira continua e complementa observando que o perfil também deve incorporar caracterís-

ticas de trabalho em colaboração apoiado em amplo processo de investigação, comunicação e criação.

Neste sentido, incorpora-se também ao perfil a descoberta de requisitos e necessidades de informação em múltiplos contextos. Com o envolvimento de diferentes personagens (usuários, clientes, parceiros de negócio, informações de mercado, redes sociais, feeds de notícias, dentre outros), o Cientista de Dados deve incrementar sua capacidade de criação e colaboração (Breitman, 2014).

Portanto, para o Cientista de Dados e Informação descobrir requisitos, envolver diferentes personagens e conduzir trabalhos em colaboração, propõe-se que ele parta de algumas orientações para nortear o trabalho etapa de *Discovery* (Oliveira; 2013):

a) A necessidade de conhecer o contexto dos dados e informações, identificando todos os elementos pertinentes (fontes, descrições e correlações);

b) A necessidade lançar visões diferenciadas para formular propostas de organização dos dados e informações;

c) A necessidade de propor categorizações para auxiliar projetos de conjuntos de dados (*clusters*);

Para desenvolver projetos em Big Data, Shiri (2014) apresenta uma abordagem que se aproxima e convalida as orientações de Oliveira, pois propõe observar o contexto segundo as facetas que auxiliam na compreensão dos elementos: Contexto/ambiente, Análise/Analytics, Metadados, Atividades e Operações, Tipos de Dados e Pessoas (Figura 3).

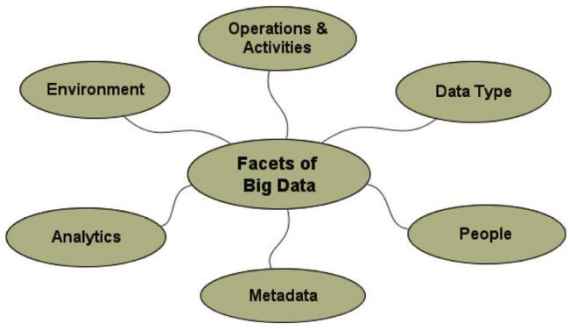


*Figura 3: Facetas utilizadas em projetos de Big Data (Shiri; 2014)*

A visão adotada por esta investigação procurou evidenciar a faceta *Data Type*. No âmbito desta faceta, Shiri continua e propõe que o processo de análise contemple também as subfacetas: *nature, context, creator, processing, publication, structure, format e access* (Figura 4).

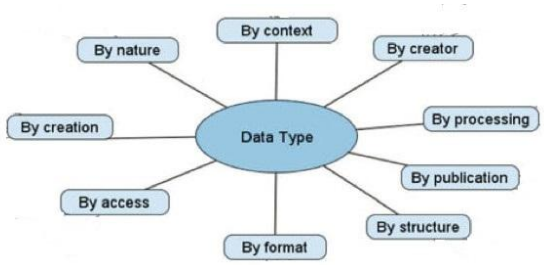


*Figura 4: Subfacetas da faceta Data Type utilizadas em projetos de Big Data (Shiri; 2014).*

## 4. Metodologia

O recorte deste estudo contempla o uso de recursos ligados a organização do conhecimento para auxiliar no desenvolvimento de projetos de Big Data, objetivando especialmente a etapa inicial de *Discovery*.

Trata-se de um estudo exploratório e tendo em vista o recorte adotado, buscou-se identificar os conjuntos de dados e de informação que possuem correlação com a temática de Acidentes de Trabalho. Esta fase foi desenvolvida, por meio de Análise da Relevância dos dados disponíveis no portal de Dados Abertos da Dataprev (10). As questões sobre a relevância da informação tratadas por Saracevic (1970; 1996b) foram utilizadas para auxiliar na definição dos conjuntos de interesse. Os itens relevantes, em geral, atenderam a questões formuladas nessa pesquisa e percebidas como de interesse para o cliente e/ou usuário. Estas questões foram desenvolvidas com o apoio de critérios, sintetizados a seguir (Ribeiro, 2012):

- As informações ou os conjuntos de dados selecionados podem ocasionar sentimentos de excitação e satisfação aos clientes e/ou usuários;
- Especificidade de pesquisas propostas para os clientes e/ou usuários estão contempladas nos conjuntos selecionados;

Oliveira complementa e corrobora a abordagem com a proposta das duas primeiras tarefas para executar o trabalho de *Discovery* na atividade de *Analytics,* são: a identificação das informações que serão necessárias para possibilitar interpretações sobre os resultados, correlações, implicações e causalidade; e determinar o(s) tipo(s) de organização para estas informações (classificações e criação de *clusters* de dados).

Após o recorte dos dados no domínio, partiu-se para a análise da faceta evidenciada e suas

subfacetas, com o intuito de auxiliar no processo para categorização dos conjuntos de dados. Conforme Shiri (2014) observa, o uso da análise de facetas possibilita delinear características, atributos e diferentes aspectos de um ambiente complexo. Ademais, em relação ao conteúdo analisado, para Li e Belkin (2008), um esquema de classificação pode refletir a possibilidade de compreender melhor os aspectos ligados ao processo de obtenção deste conteúdo.

## 5. O experimento sobre os dados de acidente de trabalho

A Previdência Social é detentora de muitos dados e informações da sociedade brasileira. Com o esforço para publicação de dados em formato aberto para poderem ser reutilizados pela sociedade (Ribeiro; Almeida, 2011), surgiram muitas possibilidades de reuso, tanto em projetos de pesquisa específicos quanto em trabalhos acadêmicos (Rodrigues, 2012; Campos; Campos; Carvalho; Lima, 2012; Germano, 2013).

Esta investigação está sendo desenvolvida no âmbito de um projeto de pesquisa em curso na Unirio e está apoiada em trabalhos anteriores desenvolvidos pelo autor. O objetivo deste estudo foi investigar um conjunto de ações que viabilizem a participação do profissional da informação em projetos de Big Data, em especial na atividade de *Analytics*. Assim, a delimitação do campo empírico se deu a partir da questão: é possível analisar os dados abertos sobre Acidentes de Trabalho disponíveis no portal da Previdência e organizá-los segundo os princípios da Organização do Conhecimento?

Neste sentido, partiu-se do pressuposto que a elaboração de modelo conceitual de dados é útil, pois permite avaliar, inclusive, a possibilidade de realizar integrações de dados e associações. Para Kent (1998), o modelo conceitual de dados nos ajuda a compreender melhor as diferentes visões da realidade. Ademais, Sukumar e Ferrel (2013) convalidam esta estratégia quando afirmam que conhecer estrutura dos dados e seus relacionamentos, pode facilitar o trabalho de análise de conjuntos de dados.

Assim, a investigação partiu da identificação dos dados e informações relevantes no Portal de Dados abertos da Dataprev e, neste sentido, foram identificados dois conjuntos de dados que poderiam ser úteis para responder a questão de partida: Acidentes de Trabalho por CBO (11) e Acidentes de Trabalho por CNAE 2.0 (12). A representação a seguir (Figura 5) foi gerada a partir de uma extensão do modelo obtido em Ribeiro e Almeida (2011).

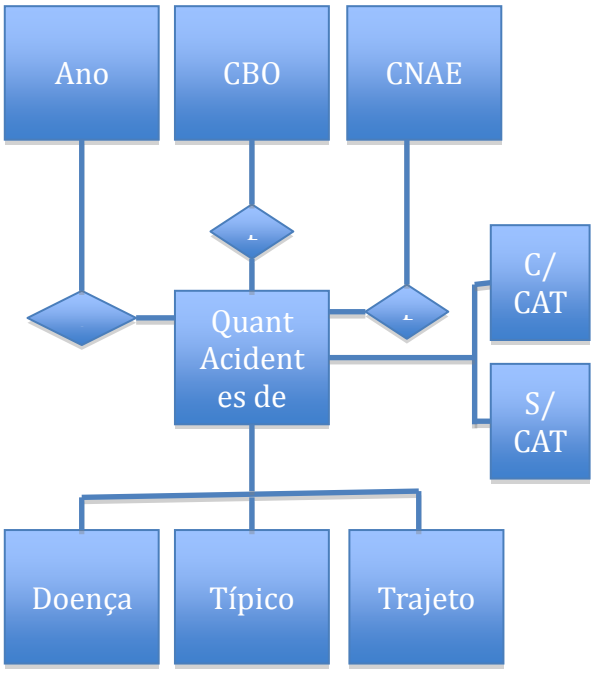


*Figura 5: Modelo de Dados dos conjuntos escolhidos (elaborado pelo autor)*

Cabe registrar que o modelo de dados utilizado permitiu representação semântica na medida em que este modelo apresenta elementos (entidades e atributos) e relações em um determinado domínio (Campos, 2001). Campos, Campos, Carvalho e Lima (2012) reafirmam esta estratégia de modelagem conceitual, pois a mesma permite uma boa forma de efetuar a representação do conhecimento em um domínio.

A partir do modelo de dados buscou-se o entendimento das dimensões propostas por Shiri. Adotou-se a faceta *Data Type* , pois o objetivo dessa pesquisa foi explorar os conjuntos de dados disponíveis. As subfacetas foram analisadas para cada conjunto de dados presente no modelo de dados, levando-se em conta a origem, o formato, o processamento e o contexto, conforme conteúdo disponível no portal da Dataprev (10) (15).

A subfaceta *context* foi incrementada com a descrição do conjunto de dados, o que permitiu uma análise mais apurada do significado do conjunto. A subfaceta *processing* foi omitida no proceso de análise, pois a forma de geração do conjunto não estava disponível no portal no momento da pesquisa.

A tabela a seguir apresenta a sistematização dessa análise:

| *Conjunto de Dados* | *Creator* | *Format* | *Context* |
|---|---|---|---|
| Quant. Acidentes-Doença | Gov/ INSS | JSON<br>XML<br>CSV | Acidentes ocasionados por qualquer tipo de doença profissional peculiar a determinado ramo de atividade constante na tabela da Previdência Social. |
| Quant. Acidentes-Típico | Gov/ INSS | JSON<br>XML<br>CSV | Acidentes decorrentes da característica da atividade profissional desempenhada pelo acidentado; |
| Quant. Acidentes-Trajeto | Gov/ INSS | JSON<br>XML<br>CSV | Acidentes ocorridos no trajeto entre a residência e o local de trabalho do segurado e vice-versa; |
| Quant. Acidentes-c/CAT | Gov/ INSS | JSON<br>XML<br>CSV | Corresponde ao número de acidentes cuja Comunicação de Acidentes do Trabalho – CAT foi cadastrada no INSS. Não são contabilizados o reinício de tratamento ou afastamento por agravamento de lesão de acidente do trabalho ou doença do trabalho, já comunicados anteriormente ao INSS. |
| Quant. Acidentes-s/CAT | Gov/ INSS | JSON<br>XML<br>CSV | Corresponde ao número de acidentes cuja Comunicação de Acidentes do Trabalho – CAT não foi cadastrada no INSS. O acidente é identificado por meio de um dos possíveis nexos: Nexo Técnico Profissional/Trabalho, Nexo Técnico Epidemiológico Previdenciário – NTEP ou Nexo Técnico por Doença Equiparada a Acidente do Trabalho. Esta identificação é feita pela nova forma de concessão de benefícios acidentários |
| CNAE | Gov/ IBGE | JSON<br>XML<br>CSV | Classificação Nacional de Atividade Econômica (99 subclasses) |
| CBO | Gov/MTE | JSON<br>XML<br>CSV | Código Brasileiro de Ocupações (192 subgrupos) |
| Ano | Gov/Dataprev | JSON<br>XML<br>CSV | Ano |

*Tabela 1: Análise das subfacetas escolhidas (elaborado pelo autor)*

A análise das subfacetas presentes na Tabela 1 apoiou a execução da etapa de *Discovery*, por meio da exploração e organização com o conteúdo levantado.

A descoberta de características e similaridades nas subfacetas permitiram a agregação e/ou sumarização dos conjuntos de dados. A identificação tanto da origem (*creator*) quanto do formato (*format*), permitem propor processos de carga no *data lake* (ver 3.2) mais ágeis, além de viabilizar discussões ligadas à gestão e ao licenciamento dos dados (13). O contexto (*context*) permite uma análise sobre os itens de dados que compõem o conjunto, possibilitando um entendimento da informação gerada. Esta dimensão também pode determinar mudança na estratégia da carga no *data lake*, na medida em que o cientista pode identificar características de similaridade na geração da informação.

Outro aspecto importante é a possibilidade de identificar correlações entre elementos. A partir do campo *context* do conjunto de dados Quant. Acidentes-Trajeto, identifica-se a possibilidade de correlacionar os locais mais perigosos em relação aos acidentes, por meio de coordenadas GPS (14) e gerando um mapa visual com recursos do *Google Maps*.

De forma análoga, o campo *context* do conjunto de dados Quant. Acidentes-Doença, indica a possibilidade de correlacionar doenças (por meio de seus códigos de doença) com informações sobre investimentos em saúde.

A dimensão do processo de geração (*processing*) foi a mais prejudicada no processo de análise, tendo em vista a inexistência de conteúdo levantado. Todavia, percebe-se que esta é uma característica fundamental para complementar a semântica dos conjuntos de dados e que poderá, em conjunto com o contexto, ser útil na definição de vocabulários específicos para a descrição dos conjuntos de dados. O uso destes vocabulários pode auxiliar na descrição a identificação da estrutura semântica dos elementos sob análise (Ribeiro; Vieira, 2014)

## 6. Considerações Finais

O uso das abordagens ligadas ao tema Big Data ainda carece de investigação pela área de Ciência da Informação. O uso dessas visões poderá auxiliar na melhoria da oferta de serviços de informação. Conhecer o ambiente de dados e informação, que deve ser originário de diferentes fontes, efetuar a organização de conjunto de dados (categorizá-los?), realizar entrevistas junto aos clientes e/ou usuários e desenvolver os modelos (tanto estruturais quanto matemáticos), contribuirá para o projeto desses serviços. Shiri (2014) convalida esta percepção quando afirma que a Ciência da Informação e em especial os métodos de organização do conhecimen-

to, são terrenos sólidos e adequados para desenvolver estudos em Big Data.

Para Minelli, Chambers e Dhiraj (2013), o momento que vivemos é especial, pois a contínua redução do custo dos equipamentos, além do uso de novos softwares e ferramentas para apoiar os processos de gestão de dados e informação, têm contribuído para um momento especial no tratamento da informação.

Quando lançamos um olhar na direção das bibliotecas e unidades de informação, é possível perceber que tanto os estoques existentes quanto as demandas por outras fontes têm crescido de forma exponencial, pois os ativos de informação de interesse para os usuários estão armazenados em diferentes bases de dados, usando bancos de dados e plataformas de computação heterogêneas (Meletiou, Katsirikov; 2009). Adicione-se a isso, a presença cada vez maior de redes sociais e aplicações de colaboração nos serviços das unidades de informação, ocasionando mudanças no comportamento dos usuários (De Jesus, Da Cunha; 2012; Santos, Da Rocha; 2012). Em suma, compreender e participar de esforços com o uso de Big Data pode ser valioso para auxiliar na estruturação de unidades de informação.

Por fim, Shiri (2014) observa que a noção de Big Data está sendo incorporada gradualmente no ambiente de informação digital, logo, é essencial que sejam realizadas investigações que contemplem o uso de outras facetas. Em especial no que diz respeito à faceta de Metadados, em decorrência de alguns aspectos emergentes como: o intenso uso de repositórios, a necessidade de criação de mecanismos para acesso eficiente, além da colaboração entre pesquisadores e instituições no ambiente de *Web of data,* bem como na necessidade de intercâmbio de informações.

Projetos de Big Data são desenvolvidos com os objetivos de criar novos produtos, compreender necessidades dos clientes e seus comportamentos, bem como perceber novos mercados. Para isto, é necessário desenvolver teorias para tratar com clientes e usuários, construindo hipóteses e identificando dados e informações relevantes. Este processo deve ser repetido e refinado, de acordo com os experimentos realizados e as respostas obtidas (Marchand; Peppard, 2013).

Espera-se que este movimento de pesquisa na área da Ciência da Informação, ilumine o caminho a ser trilhado e possibilite que outros pesquisadores interessados possam se engajar nesta discussão, levando este tema para além da tecnologia.

## Notas

(1) Neologismo criado por Alvin Tofler que junta os conceitos de produtor e consumidor.

(2) A unidade Zettabyte é equivalente, em valores aproximados, a 1.000 Exabytes, ou a 1.000.000 de Petabytes, ou ainda, a 1.000.000.000 de Terabytes.

*(3)* Tradução livre de *Cloud Computing*

(4) Também conhecida como *Apache Hadoop* é tecnologia *open source* desenvolvida pela *Google e Yahoo* para processar muitos dados em servidores, usando a noção de processamento em paralelo e uso de *clusters* (conjuntos) de computadores no processamento. Possui versões customizadas desenvolvidas por fabricantes de software e hardware.

(5) *MapReduce* é a proposta de arquitetura que deu origem à tecnologia de *Hadoop*. Usa a estratégia de dividir para conquistar, ou seja, distribui e aloca um problema muito grande em *clusters* de armazenamento, usando registros serializáveis do tipo <chave, valor>.

(6) Sigla de *Hadoop Distributed File System*. Estrutura de armazenamento de arquivos que utiliza blocos de até 64 Megabytes, que são muito menores do que os blocos de particionamento tradicionais, possibilitando a transmissão dos blocos em modalidade de *streamming*.

(7) *Machine Learning* trata o uso de algorítimos que identificam o melhor modelo para ser aplicado ao conjunto de dados.

(8) Empresas de aconselhamento estratégico é como são conhecidas as empresas de consultoria em Tecnologia da Informação, que prestam serviços de elaboração de relatórios sobre o uso e adoção de novas tecnologias.

(9) A noção de *mining* de dados passa pela extração e análise de grandes volumes de informação em busca de padrões e comportamentos.

(10) Disponível em http://dadosabertos.dataprev.gov.br

(11) Classificação Brasileira de Ocupações

(12) Classificação Nacional de Atividade Economômica

(13) O licenciamento para reuso destes conjuntos de dados precisa ser feito com a instituição provedora e deve se apoiar nos 4 tipos de licenciamento disponível: ODC-BY, ODC-ODBL, ODC-DBCL e ODC-PDDL (Open Data Commons, 2015).

(14) GPS – *Global Positioning System* – Sistema Global que identifica coordenadas geográficas das localidades.

(15) A descrição da subfaceta *context* foi extraída do Anuário Estatístico Ano 2013. Pag. 543. Disponível em http://www.previdencia.gov.br/estatisticas/

## Referencias


Abib, G. (2010). A qualidade da informação para a tomada de decisão sob a perspectiva do sensemaking: uma ampliação do campo. *Ciência da Informação*, 39(3). 73-82.

Assis, J. Moura, M. A. (2011). A Qualidade da Informação na WEB: uma abordagem semiótica. *Informação & Informação*. 16(3), 96-110.

Brietman, K. (2014). Big Data Seen from the Clouds. 2014. Palestra apresentada no 2o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/_Big_Data_Summer_School_Karin.pdf Acesso em Maio de 2015.

Campos, M. L. A. (2001). *A organização de unidades de conhecimento: o modelo conceitual como espaço comunicacional para a realização da autoria.* (Doutorado em Ciência da Informação). ECO/IBICT, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Campos, L. M.; Campos, M. L. A.; Carvalho, M. G. P.; Lima, D. V. M. (2012). Dados abertos interligados e o espaço do profissional de informação: Uma aplicação no domínio da enfermagem In: (2012) *Encontro Nacional de Pesquisa em Ciência da Informação*, Rio de Janeiro: FIOCRUZ. Disponível em: http://www.eventosecongressos.com.br/metodo/enancib2012/anais/index.php. Acesso em: Maio de 2015.

Ciarini, A. E. M. (2013). Research on Big Data and Opportunities. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro.

Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities.* Harvard Business Review Press Books.

Davenport, T. H.; Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review* 90 (10-October), 70–76.

De Jesus, D. L.; Da Cunha, M. B. (2012). Produtos e serviços da web 2.0 no setor de referência das bibliotecas. *Perspectivas em Ciência da Informação*, 17 (1), 110-133.

Dumbill, E. (2012). What is Big Data? In: (Editor) (eds). *O'Reilly Media Inc. Big Data Now: current perspectives*. O'Reilly Media:California. 2012. Disponível em: http://www.oreilly.com/data/free/files/big-data-now-2012.pdf. Acesso em: Maio 2015.

EMC. Brazil country brief. (2014). *The Digital Universe of opportunities.* Disponível em: http://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014-brazil.pdf. Acesso em: Maio 2015.

Ferreira, A. M. J. F. C.; Santana, R. C. G.; Vidotti, S. A. B. G. (2012). Second life: perspectivas para potencializar o acesso a dados públicos. In: (2012). *Encontro Nacional de Pesquisa em Ciência da Informação*, Rio de Janeiro: FIOCRUZ. Disponível em: http://www.eventosecongressos.com.br/metodo/enancib2012/anais/index.php. Acesso em: Maio de 2015.

Florissi, P. (2014). Big Data: Challenges and Opportunities. Palestra apresentada no 2o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro.

Florissi, P. EMC On Big Data. (2012) . Disponível em: https://www.carecorenational.com/healthcaresummit/powerpoints/PatriciaFlorissiPhD.pdf. Acesso em: Maio de 2015.

Fonseca, M. J.; Gonçalves, M. A.; Oliveira M. O. R.; Tinoco, M. A. C. (2008). Tendências sobre as comunidades virtuais da perspectiva dos prosumers. *RAE Eltrônica*. 7 (2). Disponível em http://rae.fgv.br/rae-eletronica/vol7-num2-2008/tendencias-sobre-comunidades-virtuais-perspectiva-prosumers. Acesso em: Maio de 2015.

Fox, P.; Hendler, J. (2011). Changing the Equation on Scientific Data Visualization. Science 331, 705. Disponível em: <http://data2discovery.org/dev/wp-content/uploads/2013/05/Fox-and-Hendler_Visualization_Science-2011-Fox-705-8.pdf>. Acesso em: Maio de 2015

Germano, E. C. (2013). *Modelos de negócios adotados para o uso de dados governamentais abertos: estudo exploratório de prestadores de serviços na cadeia de valor dos dados governamentais abertos.* 2013. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, University of São Paulo, São Paulo, 2013. Disponível em: http://www.teses.usp.br/teses/disponiveis/12/12139/tde-10012014-155226/. Acesso em: Maio de 2015.

Goguen, J. A. (1994). Requirements Engineering as the reconciliation of social and technical issues. In: *Jirotka, M. e Goguen, J. A. (1994) (Ed.). Requirements Engineering - Social and Technical Issues*. London, 1994. p.165-199.

Gopintah, M. A.; Das, P. (1997). Classification and representation of Knowledge. *Library Science with a Slant to Documentation and Information Studies*. 34 (2). 85-90.

Hashem et. al. The rise of "Big Data" on cloud computing: Review and open research issues. *Information Systems*, 47 (2015) 98-115

Heath T..; Bizer C. (2011). *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers.

HjØrland, B.; Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6), p. 400-425.

Ibm (2015). IBM BigData e Analytics Hub. *Why only one of the 5 Vs of big data really mattersDisponível em:* http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters. Acesso em: Maio 2015.

InfoExame (2014). Solução de Big Data é um dos segredos da seleção alemã na copa. Disponível em : http://info.abril.com.br/noticias/it-solutions/2014/07/solucao-de-big-data-e-um-dos-segredos-da-alemanha-na-copa-do-mundo.shtml. Acesso em: Maio de 2015.

Li, Y.; Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management* .44 .1822–1837.

Khan, N et. al. (2014). Big data: survey, technologies, opportunities, and challenges.(Report). *The Scientific World Journal*, 14.

Kent, W. (1998). *Data and Reality*. 1stBooksLibrary. Bloomington.

Mann, S.; Nolan, J; Wellman, V. Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments. Surveillance & Society 1:3 (2003) 331-355

Marchand, D. A.; Peppard, J. (2013). Why IT Fumbles Analytics*. Harvard Business Review*. Jan-Fev.

Mattoso, A. Scientific Workflows and Big Data. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro..

Meletiou, A; Katsirikou, A. (2009). Methodology of analysis and interrelation of data about quality indexes of library services by using data-and knowledge-mining techniques. *Library Management* . 30 (3), 138-147.

Mello, R., Leite, L. R., e Martins, R. A. (2014). Is Big Data the Next Big Thing in Performance Measurement Sys-

tems? Guan, Y e Liao, H. (eds). *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*.

Minelli, M.; Chambers, M.; Dhiraj, A. (2013). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. Wiley CIO Series.

Oliveira, A. (2013). Data Science and Data Analytics. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro.

ONUBR. (2013). ONU: Dos 7 bilhões de habitantes do mundo 22 de março de 2013. Disponível em http://www.onu.org.br/onu-dos-7-bilhoes-de-habitantes-do-mundo-6-bi-tem-celulares-mas-25-bi-nao-tem-banheiros/ Acesso em: Maio de 2015.

Open Data Commons (2015). Disponível em: http:// www.opendatacommons.org/about. Acesso em : junho de 2015.

Porto, F. (2013). Big Data in Astronomy: The LIneA-DEXL case. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro.

Power, D. J. (2014). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, 3 (2).

Ribeiro, C. J. S. (2008). *Diretrizes para o projeto de portais de informação: uma proposta interdisciplinar baseada na Análise de Domínio e Arquitetura da Informação.* 298 f. Tese (Doutorado em Ciência da Informação) – Convênio UFF/IBICT, Rio de Janeiro.

Ribeiro, C. J. S. (2012). Entendimento de requisitos de sistema com abordagem orientada ao domínio. *DataGramaZero - Revista de Informação*. (Abril 2012). 13 (2). Disponível em http://www.dgz.org.br/abr12/Art_01.htm. Acessado em: Maio de 2015.

Ribeiro, C. J. S.; Almeida, R. F. (2011). Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade. In*: Elmira Simeão, Jorge Henrique Cabral Fernandes, Isa Maria Freire. (Org.). XII Enancib - Políticas de Informação para a Sociedade* Brasília: Thesaurus.

Ribeiro, C. J. S.; Pereira, D. V. (2014). El proceso de publicación e los datos gubernamentales abiertos acerca de la clase de la Seguridad Social Brasileña de Vocabulario Controlado del Gobierno Electrónico (VCGE). In: *Vanderkast E. J. S. (ed) (2014). El acceso a la información gubernamental : experiencias y expectativas.* Cidade do México: UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información

Robredo, J. (2011). Do documento impresso à informação nas nuvens: reflexões. *Liinc em Revista,* 7 (1), 19-42. Disponível em: http://www.ibict.br/liinc . Acesso em maio de 2015.

Rodrigues, F. A. (2012). *Mapeamento de tecnologias informacionais sobre dados abertos em saúde pública: destino de repasses financeiros federais*. Dissertação (Mestrado em Ciência da Informação). Universidade Estadual Paulista. Faculdade de Filosofia e Ciências.

Santos, E. L.; Da Rocha, S. M. (2012). O blog como ferramenta de comunicação entre a biblioteca e seus usuários: a experiência da biblioteca Lydio Bandeira de Mello, da Faculdade de Direito da Universidade Federal de Minas Gerais. *Encontros Bibli*, 17 (33). 134-152.

Santos, I. H. R. (2014). BigData Research and Developement at Petrobras. 13 de maio de 2014. Palestra apresentada no 2o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em: <http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Ismael_BigDataTOOL_SummerSchool_v2.pdf>. Acesso em: Maio de 2015.

Sarecivic, T. (1970). The concept of "relevance" in Information Science: an historical review. In: *Saracevic, T (Ed.) Introduction to Information Science.* New York: R. R. Bowker, 111-154.

Saracevic, T. (1996a). Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação,* 1 (1). 41-62,.

Saracevic, T. (1996b). Relevance reconsidered 1996. Information Science: Integration in Perspectives. *In: International Conference of Library and Information Science (COLIS, 2).* Copenhagen, Denmark, 14-17 Oct.

Sathi, A. (2013). *Big Data Analytics: Disruptive Technologies for Changing the Game*. Mc Press. 2013.

Seymour, C. (2014). The State of Big Data. *EContentMag.com*. Jan-Feb. 26-27.

Siegel, E. (2013). Predictive Analytics. Wiley. New Jersey.

Sukumar, S. R.; Ferrel, R. K. (2013). 'Big Data' collaboration: Eploring, recording, and sharing enterprise knowledge. *Information Services & Use*. 33. 2013.

Tavares, E. BIG DATA in Business. 12 de maio de 2014. Palestra apresentada no 2o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em: http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao_Elaine_Tavares.pdf . Acesso em: Maio de 2015.

Tofler, A. *Third Wave: The classic study for tomorrow.* Bantam Books: New York. 1984

Xie, C.; Bagozzi, R.; Troye, S. (2008). Trying to prosume: toward a theory of consumers as co- creators of value. *Journal of the Academy of Marketing Science,* 36, 62-78.

Wurman, R. S. *Ansiedade de Informação 2.* São Paulo: Editora de Cultura, 2005.

Zikopoulos, P., deRoos, D., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D. (2013). Har- ness the Power of Big Data: *The IBM Big Data Platform*. New York: McGraw Hill.