

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
NÚCLEO DE INFORMAÇÃO TECNOLÓGICA E GERENCIAL
CURSO DE ESPECIALIZAÇÃO EM GESTÃO DE INFORMAÇÃO E
PESSOAS**

Pedro Augusto Moreira Dias

**APLICAÇÃO DE *BIG DATA* AO SETOR PÚBLICO: ESTUDO DE CASO DA
PLATAFORMA DATAVIVA**

Belo Horizonte
2017

Pedro Augusto Moreira Dias

**APLICAÇÃO DE *BIG DATA* AO SETOR PÚBLICO: ESTUDO DE CASO DA
PLATAFORMA DATAVIVA**

Trabalho de Conclusão de Curso
apresentado ao Curso de
Especialização em Gestão de
Informação e Pessoas da Escola de
Ciência da Informação da
Universidade Federal de Minas
Gerais

Orientador: Prof. Dr. Jorge Tadeu
Ramos Neves – UFMG

Belo Horizonte
Escola de Ciência da Informação - UFMG
2017

RESUMO

Surge, há aproximadamente uma década, (e populariza-se muito rapidamente) o termo “*Big Data*”, que se refere, em tradução literal, ao grande volume de dados – estruturados e não estruturados – que podem impactar enormemente as relações, os processos e os negócios das organizações, de acordo com diversos autores e estudos. Apesar da crescente popularidade, o termo ainda é bastante mal-compreendido, não havendo consensos quanto ao que realmente vêm a ser *Big Data* e quais seriam as tecnologias fundamentais que o suportam. Este trabalho busca apresentar uma iniciativa (na forma de um caso) de *Big Data* desenvolvida pelo Governo de Minas Gerais: a plataforma DataViva. O seu modelo de desenvolvimento, a partir de um processo de Governança Colaborativa, se assemelha ao chamado “Ecossistema de Dados”, que será apresentado também como potencial facilitador a iniciativas de *Big Data*. Busca também elucidar o envolvimento governamental na gestão ativa (e transparente) de dados e informações. Apresenta conceitos relacionados ao *Big Data*, Tomada de Decisão, Qualidade de Dados, Open Data, Cloud Computing, e aspectos gerais e específicos do Setor Público, além do caso de Minas Gerais, conforme contexto específico que culminou no Projeto DataViva, que será apresentado como referência ao *Big Data* na forma de um caso específico. Ao final da pesquisa, são contextualizadas as considerações finais e apontamentos acerca do estudo de caso apresentado, bem como sobre iniciativas de *Big Data* de modo geral e algumas de suas potencialidades e desafios.

Palavras-chave: Dados, DataViva, Big Data, Ecossistema de Dados, Open Data, Setor Público, Informação, Tomada de Decisão, Governança Colaborativa, Formulação de Políticas Públicas.

ABSTRACT

The term "*Big Data*", which refers in literal terms to the large amount of data - structured and unstructured - has been found to have a major impact on relations, processes and the business of organizations. Despite its growing popularity, the term is still misunderstood, with no consensus as to what Big Data really is and the underlying technologies that support it. This paper aims to present a Big Data initiative developed by the Government of Minas Gerais: the DataViva platform. Its development model, following a process of Collaborative Governance, is similar to the so-called "Data Ecosystem", which will be presented as a potential facilitator for similar initiatives, dealing with government involvement in the active and transparent management of data and information. It presents concepts related to Big Data, Decision Making, Data Quality, Open Data, Cloud Computing, and general aspects of the Public Sector and the case of Minas Gerais according to the specific context that culminated in the DataViva Project. At the end of the research we presented the final considerations about the case study as well as about *Big Data* initiatives are contextualized some of their potentialities and challenges.

Key words: Data, DataViva, Big Data, Open Data, Data Ecosystem, Public Sector, Information, Decision Making, Public Policy Formulation, Collaborative Governance.

“A ‘explosão da informação’, sobre a qual muito se comenta e se escreve, é também, em grande medida, a explosão da informação errada ou mal-organizada.”

(Murray Gell-Mann)

SUMÁRIO

1 INTRODUÇÃO	7
2 OBJETIVOS	12
3 REFERENCIAL TEÓRICO	13
3.1 Aspectos iniciais: acerca de dados, informações e conhecimentos	15
3.1.1 Qualidade de dados	18
3.2 Open Data – Dados Governamentais abertos.....	21
3.3 – <i>Cloud Computing</i> : Computação em Nuvem.....	23
3.5 <i>Big Data</i>	26
3.6 “Ecosistema de dados”	29
3.7 Aspectos do Setor Público e o caso de Minas Gerais	32
4 METODOLOGIA	38
5 ESTUDO DE CASO: Plataforma DataVIVA	40
6 CONSIDERAÇÕES FINAIS	Erro! Indicador não definido.
Referências	60
ANEXO I:	62
ANEXO 2: TEXTO TINNO COMUNICAÇÃO.....	63

1 INTRODUÇÃO

O período compreendido entre o final do século XX e o início do século XXI foi marcado por uma enorme revolução das Tecnologias da Informação e Comunicação (TIC's), com profundas implicações socioeconômicas. A partir (I) dos novos modelos e instrumentos de interação entre indivíduos e organizações, (II) do apogeu de novos movimentos sociais e culturais, (III) das novas economias, informacional e global, e (IV) dos conceitos de sociedades em redes, pode se dizer que todas as relações produtivas, de poder, e de experiências, sem exceção, foram redefinidas na sociedade contemporânea. (DE PAULA, 2008)

Sob o ponto de vista dos indivíduos inseridos neste contexto atual, percebemos a grande difusão informacional de nossos tempos, caracterizada por uma vasta quantidade e variedade de informações disponíveis e acessíveis, difundidas através de diversos tipos de recursos tecnológicos, que nos conectam e nos colocam diante de inúmeras realidades, acontecimentos, fatos, e informações em tempo real, sem restrições. Grandes volumes de dados e informações são gerados e transmitidos, todos os dias, pelas inúmeras interações de bilhões de pessoas que utilizam telefones celulares e *smartphones*, computadores, *tablets*, aparelhos portáteis, e tantos outros dispositivos eletrônicos. (CAMPOS, 2015)

Até mesmo os custos relacionados ao armazenamento da informação, que antes limitavam a sua produção, difusão e arquivamento de publicações, hoje caem rapidamente e não constituem um fator limitante para a explosão de dados produzidos diariamente pelas novas tecnologias e seus usuários, seja de maneira voluntária (consciente) ou não. Não seria de todo equívoco ponderar que já, há muitos anos, vivemos em uma era de “excesso de informação”, já que a quantidade de dados e informação disponível e acessível excede em muito as nossas capacidades de obtenção, seleção, processamento e assimilação, não só a nível individual, mas também na sociedade e nas organizações.

Além disso, no contexto global, de elevada disputa e competitividade econômica, diversos países, subdesenvolvidos, emergentes, ou mesmo “desenvolvidos”, dentre eles, o Brasil, encontram-se em condições de recessão econômica, ou como muito se tem dito na atualidade, em crise. Sabe-se, todavia, que não existem truques, fórmulas mágicas ou mesmo políticas públicas pré-estabelecidas (e miraculosas), que façam as instituições passar ilesas por um contexto de instabilidade econômica. Entende-se que, a partir de um breve olhar sobre todo este contexto, somente a partir de estratégias, investimentos (públicos e privados) e políticas de longo prazo e bem definidas seria possível sobressair às dificuldades impostas por contextos de crise, e, a partir daí, retomar o desenvolvimento.

A nova economia do séc. XXI, global e informacional, traz consigo o fenômeno (e os desafios) cada vez maior(es) da busca de informações internas ou externas às organizações para referendar a tomada de decisões. Para atender a essas demandas, surgem, a todo o momento, novas tecnologias ligadas à inteligência, direcionada especificamente para a tomada de decisões. (TAURION, 2012)

Segundo o relatório “*Big Data, Big Impact: New Possibilities for International Development*” do Fórum Econômico Mundial, não apenas as empresas e organizações privadas, mas também pesquisadores e formuladores de políticas públicas estão começando a perceber o grande potencial e as oportunidades de se canalizar *grandes volumes de dados* em informações úteis e acionáveis, que possam ser utilizadas para identificar as necessidades das pessoas, aprimorar a prestação de serviços, e prever e prevenir crises e eventos adversos para o benefício de toda sociedade. (WORLD ECONOMIC FORUM, 2012)

Em meio a todo este contexto, surge há aproximadamente uma década (e populariza-se rapidamente) o termo “*Big Data*”, que se refere, em tradução literal, ao grande volume de dados – estruturados e não estruturados – que impactam as relações e o negócio das organizações no dia a dia. Apesar da popularidade, o autor destaca que o termo ainda é bastante mal compreendido,

não havendo consensos quanto ao que realmente vêm a ser *Big Data*, bem como os preceitos e tecnologias fundamentais que o suportam. Existem, ainda, muitas dúvidas sobre como extrapolar o conceito, ou seja, como sair do discurso conceitual e criar soluções tecnológicas efetivas que agreguem valor para as organizações. (TAURION, 2012)

Ilustrando ainda esta tendência, de acordo com o relatório do World Economic Forum, o governo dos Estados Unidos anunciou, em 2012, o “*Big Data Research and Development Initiative*”, um programa que reúne diversos departamentos e agências federais e totaliza investimentos da ordem de 200 milhões de dólares para desenvolver e aprimorar ferramentas e técnicas necessárias para capturar, acessar, organizar e extrair informações de grandes volumes de dados. O objetivo da iniciativa é acelerar o ritmo de descoberta em ciência e engenharia de dados, fortalecer a segurança nacional e transformar o ensino e a aprendizagem através da extração de conhecimento dos bancos de dados. (WORLD ECONOMIC FORUM, 2012)

Há, contudo, bastante tempo que recursos de Tecnologias da Informação já contemplam aplicações de *Data Mining* (mineração de dados), *Business Intelligence* e *CRM (Customer Relationship Management)*, por exemplo, para tratar justamente de questões como a análise de dados, tomadas de decisões e outros aspectos relacionados ao negócio (TAURION, 2012).

Grosso modo, e, conforme será exposto no decorrer do trabalho, a ideia central do *Big Data* pode ser compreendida como a análise de grandes volumes de dados para a geração de resultados e *insights* importantes que, em volumes menores ou sem determinadas tecnologias específicas, talvez não pudessem ser alcançados. Neste sentido, a proposta do *Big Data* é a de oferecer uma abordagem *ainda mais* ampla no tratamento do aspecto (cada vez mais) caótico dos dados, buscando tornar as suas referidas aplicações, e também todas as demais, mais eficientes e precisas. (CAMPOS, 2015)

Tendo em vista este fenômeno moderno e as suas implicações para o campo da Ciência da Informação, o tema proposto para o presente projeto de

pesquisa será “Aplicações de *Big Data* ao Setor Público: ferramentas de apoio à tomada de decisão e formulação de políticas públicas”, e ele buscará elucidar conceitos e aplicações relativas ao *Big Data*, destacando potencialidades e alguns dos desafios para o desenvolvimento e utilização dessas soluções pelo Poder Público, tendo em vista as suas especificidades e a rápida expansão da oferta e demanda de informações dentro e fora dos ambientes organizacionais.

Deste modo, o presente estudo buscará responder ao seguinte problema de pesquisa: “existem aplicações, caracterizadas como *Big Data*, passíveis de implementação pelo Poder Público, que constituam ferramentas de apoio a Tomada de Decisão e formulação de Políticas Públicas?”, através da apresentação de conceitos relacionados ao *Big Data* e de um *estudo de caso* correlato de uma plataforma de dados – O DataViva - delineando algumas das diversas potencialidades e desafios na área, assim como aplicações dos conceitos abordados e referenciados pelo estudo.

A plataforma começou a ser desenvolvida em 2011 pelo Governo de Minas Gerais, por meio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG, uma instituição pública de fomento à pesquisa e inovação. Em sua concepção, a plataforma DataViva contou com um modelo de governança colaborativa bem definido, possível apenas a partir de uma parceria entre diversos órgãos e entidades do Estado, ministérios, entidades privadas e o laboratório de pesquisa MIT Media Lab, do Massachusetts Institute of Technology – MIT.

Espera-se, através da apresentação do caso da plataforma DataViva, demonstrar que, mais que um modismo, é factível a aplicabilidade de soluções e iniciativas de *Big Data* tanto em organizações do setor privado quanto em organizações públicas. Será apresentando também o modelo de governança colaborativa adotado no desenvolvimento do DataViva, corroborando uma visão de Governos como “catalisadores de dados e informações”, e destacando também possibilidades de parceria com os setores privado e acadêmico para desenvolvimento de soluções que aprimorem a tomada de decisão e a

formulação de políticas públicas, buscando apontar potencialidades e desafios no desenvolvimento e uso dessas ferramentas.

2 OBJETIVOS

2.1 Objetivo geral

Verificar a aplicabilidade de soluções tecnológicas e informacionais de *Big Data* ao setor público, a partir do estudo de caso da Plataforma DataViva, destacando potencialidades e desafios para implantação de iniciativas similares.

2.2 Objetivos específicos

- Conceituar *Big Data*;
- Avaliar aspectos relacionados à aplicação de *Big Data* como ferramenta de apoio à Tomada de Decisão em âmbito governamental;
- Investigar iniciativa ou case de *Big Data* em uma organização do setor público;
(case selecionado: Plataforma DataVIVA – Governo de Minas Gerais, onde trabalho como servidor público em cargo efetivo de especialista em políticas públicas e gestão governamental);
- Destacar potencialidades e desafios identificados na experiência mineira, que possam ser avaliados para o desenvolvimento de iniciativas de *Big Data*.

3 REFERENCIAL TEÓRICO

Dados e informações constituem a base para quase todos os processos e tomadas de decisão em âmbito organizacional. Entretanto, os dados corporativos não são auto-gerenciáveis. Com o passar do tempo, o volume de dados e informações que vai sendo acumulado e processado pelas organizações exige constantes manutenções. É necessário gerir esses dados de forma a alinhar pessoas, processos, regras de negócio e tecnologias com a própria estratégia e demandas da organização. (SIDNEY, 2015)

A utilização inteligente (e por sua vez, estratégica) dos dados pelas organizações é um tema relativamente recente em discussão dentro do estudo das Tecnologias e Ciências da Informação e Comunicação. Cada vez mais as instituições percebem e entendem os seus dados e informações como recursos valiosos. E, como com qualquer ativo valioso, elas também percebem a premente necessidade de se gerenciá-los adequadamente. Por causa disso, é de suma importância que esses ativos informacionais sejam seguros, de alta qualidade, e que estejam sempre disponíveis, caso contrário, perdem seu valor para as organizações. (SIDNEY, 2015)

Além disso, observa-se a importância, e marcante característica contemporânea, da constante busca por se dispor de informações corretas e íntegras no menor tempo possível, no âmbito das organizações. A qualidade dos dados e informações torna-se cada vez mais um aspecto fundamental para garantir a competitividade das empresas e a eficiência dos órgãos públicos, provendo maior agilidade e suporte aos processos decisórios e permitindo uma rápida adequação diante de contextos de mudanças socioeconômicas. (SANTOS, 2010, apud SIDNEY, 2015)

Para Davenport (1998) contudo, o que está em falta nas organizações não é a informação em si, já que essa se encontra disponível em todo lugar. Faltam, na verdade, ferramentas tecnológicas e mecanismos de gestão para entregar a informação correta (e necessária) de forma tempestiva e com alto grau de qualidade, segurança e confiabilidade, ou seja, falta um eficiente

gerenciamento de informações, a nível organizacional. (DAVENPORT, 1998, apud MAFLI, 2015).

Além disso, a constante transformação tecnológica contemporânea evidencia a existência de uma sociedade informacional, que demanda respostas rápidas, precisas e abrangentes, evidenciando a chamada “explosão” da informação, que demandará (cada vez mais) às organizações, públicas ou privadas, inovação e adaptação a este novo contexto, dinâmico e imprevisível (DE PAULA, 2008).

Choo (2003), por sua vez, destaca que "a informação é um componente *intrínseco* de tudo que uma organização faz", ou seja: para toda e qualquer organização, e em qualquer atividade econômica, científica, ou mesmo governamental, a informação (e o conhecimento, conseqüentemente) é um fator inerente e essencial em todos os seus processos e resultados. O autor afirma que:

“sem uma clara compreensão dos processos, organizacionais e humanos, pelos quais as informações se transformam em percepções, conhecimentos e ações, as organizações não são capazes de perceber plenamente a importância das suas fontes (e tecnologias) de informação”. (CHOO, 2003, apud MAFLI, 2015, p.16 – grifo nosso).

Existe, ainda de acordo com o autor, uma relação indissociável entre a informação (e o conhecimento) e a tomada de decisões: para tomar-se uma decisão, deve-se primeiramente elencar todas alternativas, ou cenários, possíveis e identificar todas as suas potenciais conseqüências, considerando-se a tomada de uma decisão racional. Para que isso seja possível de fato, entretanto, é necessário verificar todos os dados, informações e conhecimentos disponíveis (e necessários), dentro e fora da organização, e a partir deles, extrair as informações essenciais para a solução de problemas específicos (CHOO, 2003, apud MAFLI, 2015).

É inegável, contudo, que essa verificação de todos os dados, informações e conhecimentos disponíveis e necessários, interna e externamente à organização, constitui-se como um enorme desafio, visto que

estes nem sempre estão disponíveis, acessíveis ou mesmo estruturados. (CANARY, 2013)

De acordo com Ladeira (2009), a expressiva utilização de recursos informacionais nas organizações contemporâneas, e a evolução dos sistemas de informação, têm como principais desafios: (I) a busca ou localização das informações, (II) a seleção ou filtragem destas, (III) a sua organização e, por fim, (IV) a disponibilização da informação *de valor* em tempo oportuno, o que é cada vez mais aprimorado à medida que novas estratégias, tecnologias e ferramentas são desenvolvidas. (LADEIRA, 2009 apud MAFLI, 2015).

Além disso, diante do imenso rol de dados e informações gerados diariamente dentro e fora das organizações, é necessário gerenciar adequadamente os ativos informacionais para o alcance dos resultados desejados. Para se chegar a uma tomada de decisão realmente efetiva, é preciso garantir a integridade, precisão, atualização, interpretabilidade e valor da informação em que se baseia, que consiste, em última análise, em um recurso estratégico sob a perspectiva da obtenção de vantagem competitiva. (OLIVEIRA et al., 2014 apud SIDNEY, 2015).

Tratando-se, antes, contudo, de conceitos e aspectos iniciais para as discussões do presente trabalho, é necessário destacar a diferença entre dados, informação e conhecimento, conceitos essencialmente correlatos, porém distintos entre si.

Embora alguns autores (DATE, 2004; ELMASRI, 2005) considerem que dados e informações podem ser compreendidos como sinônimos, a maior parte da literatura aponta para significativas diferenças entre estes dois conceitos. O conhecimento, por sua vez, diferencia-se também dos demais, mostrando-se um conceito mais amplo, mais complexo e dinâmico, e ao mesmo tempo, mais desafiador, em termos de sua gestão eficaz. (SIDNEY, 2015)

3.1 Aspectos iniciais: acerca de dados, informações e conhecimentos

Davenport e Prusak (1998) definem dados como um conjunto de “fatos ou registros objetivos e discretos sobre eventos”. Acrescenta-se a isso, que, em um contexto organizacional, seria possível descrever dados como registros estruturados, oriundos ou resultantes das transações e processos organizacionais. (DAVENPORT; PRUSAK, 1998, apud CANARY, 2013, p. 9)

Segundo os autores, dados se diferenciam de informações por constituírem-se como observações simples sobre o estado do mundo, sendo (mais) facilmente estruturados (que a informação em si), passíveis de serem obtidos por máquinas e sistemas, sendo também quantificáveis e transferíveis. (DAVENPORT; PRUSAK, 1998, apud CANARY, 2013)

Segundo Rezende: “os dados representam a matéria-prima a ser utilizada na produção de informações.” (REZENDE, 2003, apud SIDNEY, 2015 p.14).

Sob essa mesma perspectiva, é comum encontrar outros autores que também vislumbram os dados como sendo matéria prima ou base para a criação de informação e conhecimento. Para se transformarem em informação, os dados precisam ser processados, ou seja, os dados devem ser classificados, armazenados e relacionados para ganhar intencionalidade e significado e, então, serem capazes de informar algo (CHIAVENATO, 2008, apud SIDNEY, 2015).

Os dados consistiriam, assim, na matéria bruta, e, portanto primária, essencial para a criação da informação e fundamentando, posteriormente, a tomada de decisão. Dados são convertidos em informação apenas quando algum significado, valor ou contexto é adicionado a eles. (DAVENPORT, PRUSAK, 1998, apud MAFLI, 2015)

De acordo com Zaidan (2008):

“Entende-se por dado um padrão, a menor unidade possível. Uma letra, número ou dígitos, que isoladamente não tem nenhum significado claro. [...] a informação é o dado trabalhado, tratado, inserido num contexto. É um conjunto de dados que tem algum significado.”
(ZAIDAN, 2008, apud SIDNEY, 2015 p. 18)

O dado simples, isolado de contexto, não possui, por si só, valor ou significado para a organização; ele descreve apenas uma parte componente do todo contextual, sem prover ou ser constituído, isoladamente, de julgamento ou interpretação. Por outro lado, informação constitui-se como dados dotados de relevância e propósito, requerendo unidades específicas de análise, exigindo consensos em relação a seu significado e exigindo certo grau mínimo de mediação humana no seu uso e tratamento. (DAVENPORT, PRUSAK, apud MAFLI, 2015)

Figura 01 – Modelo de evolução de dados, informações e do conhecimento



Fonte: Canary, 2013, pg.12

Popularmente, dados também podem ser considerados apenas índices ou registros pontuais, contudo, isso não significa que os dados não sejam fundamentais para as organizações, pelo contrário – dados compõem e dão suporte à praticamente todas as principais rotinas organizacionais. (SIDNEY, 2015)

Quadro 01 – Conceitos e características: Dados, Informação e Conhecimento

Dados	Informação	Conhecimento
<p>Simple observações sobre o estado do mundo:</p> <ul style="list-style-type: none"> ✓ Facilmente estruturados; ✓ Facilmente obtidos por máquinas; ✓ Frequentemente quantificados; ✓ Facilmente transferíveis. 	<p>Dados dotados de relevância e propósito:</p> <ul style="list-style-type: none"> ✓ Requer unidade de análise; ✓ Exige consenso em relação ao significado; ✓ Exige necessariamente a mediação humana. 	<p>Informação valiosa da mente humana. Inclui reflexão, síntese; contexto:</p> <ul style="list-style-type: none"> ✓ De difícil estruturação; ✓ De difícil captura em máquinas; ✓ Frequentemente tácito; ✓ De difícil transferência.

Fonte: Davenport (1998) apud Canary, 2013 p.23

O quadro acima sintetiza algumas características e conceitos essenciais referentes às distinções entre dados, informação e conhecimento, de acordo com os critérios de Davenport, 1998.

“[...] (conhecimento, por sua vez) é uma mistura fluida de experiência condensada, valores, informação contextual e insight experimentado, a qual proporciona uma estrutura para a avaliação e incorporação de novas experiências e informações. Nas organizações, ele está embutido não só em documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais. Ele tem origem e é aplicado na mente dos conhecedores.” (DAVENPORT & PRUSAK, 1998, apud CANARY, 2013 p.22)

Muitas vezes, contudo, as organizações tendem a acumular uma quantidade desnecessária de bases de dados e informações, baseadas em uma perspectiva equivocada que depreende que, se uma organização armazenar enormes coleções e conjuntos de dados, conseqüentemente conseguirá acumular maior conhecimento, e, portanto, tomar decisões mais acuradas, objetivas e corretas. (SIDNEY, 2015)

Entretanto, o acúmulo de dados e informações, por si só, no âmbito das organizações, cria apenas a ilusão da acuidade científica nos processos decisórios e de gestão e rotinas da informação. O excesso de dados pode, e muito, dificultar a compreensão e a identificação de dados significativos em meio aos demais. (SIDNEY, 2015)

3.1.1 Qualidade de dados

A qualidade de dados começa a surgir expressivamente nos estudos acadêmicos a partir do início dos anos 1990. Entretanto, o expressivo reconhecimento da sua importância nas organizações é um fenômeno muito mais recente. (SIDNEY, 2015)

Segundo os autores do Data Management Body of Knowledge - DAMA DMBOK, o gerenciamento dos dados constitui-se nas atividades de desenvolvimento, execução e supervisão de planos, políticas, programas e práticas que controlam, protegem, entregam e aumentam o valor dos ativos de dados e de informação nas organizações (DAMA, 2009, apud BARBIERI,

2012). Entretanto, o gerenciamento efetivo, de dados e informações, tem diante de si alguns desafios relacionados principalmente a razões como:

- Aumento exponencial da quantidade de dados com o passar do tempo;
- Dispersão dos dados nas organizações, já que são coletados por muitos indivíduos que utilizam diferentes métodos e dispositivos;
- Origem dos dados, que podem vir de fontes internas, externas e até mesmo pessoais; (BARBIERI, 2012, p.8)

Para além do gerenciamento de dados, Santos (2010), define que a Governança de Dados pode ser conceituada de forma mais ampla como a visão estratégica a ser adotada pela organização em relação aos seus dados. Segundo o autor: “Governança, de modo geral, refere-se à forma como a organização age para garantir que as estratégias sejam implementadas, monitoradas e alcançadas”. Além disso, a Governança de Dados asseguraria que os dados tenham confiabilidade, aspecto tanto crítico como fundamental no âmbito da qualidade dos dados e das informações. (Santos, 2010, apud SIDNEY, 2015 p.18).

Nesta linha, o gerenciamento e a governança de dados e informações são importantes funções para as organizações da era da informação à medida que dados e informações em si, são cada vez mais reconhecidos como valiosos ativos das organizações. (BARBIERI, 2012)

Quadro 02 – Critérios ou dimensões importantes da Qualidade de Dados

Categoria	Critérios ou dimensões de qualidade de dados
Intrinseca	Acuracidade: Dados corretos e confiáveis; Objetividade: Os dados são imparciais; Confiabilidade: Os dados são verdadeiros; Reputação: Fonte e conteúdo dos dados são confiáveis;
Acessibilidade	Acessibilidade: Os dados estão disponíveis; Segurança do acesso: O acesso aos dados é restrito;
Contextual	Relevância: Os dados são aplicáveis e úteis; Valor agregado: Os dados são benéficos e proporcionam vantagens; Temporalidade: Os dados estão atualizados; Completeza: Os dados não estão extraviados e são suficientes; Quantidade apropriada: Volume apropriado de dados;
Representacional	Interpretabilidade: Os dados estão em linguagem apropriada; Facilidade de entendimento: Os dados são facilmente compreendidos Representação concisa: Os dados estão representados de forma compacta Representação consistente: Dados apresentados num mesmo formato

Fonte: SIDNEY, 2015 pg. 28

O quadro acima, adaptado de Santos (2010), Xu et al (2002), e Wang e Strong (1996), traz uma síntese elaborada por Sidney (2015), dos critérios e dimensões mais relevantes da qualidade de dados, segundo a bibliografia referenciada pela autora. De acordo com ela, existem 15 (quinze) dimensões da qualidade dos dados que perpassam as categorias de aspectos intrínsecos, critérios de acessibilidade, critérios contextuais e por fim, representacionais. No âmbito de cada critério, é apresentado o parâmetro de qualidade considerado adequado de acordo com os autores. (SIDNEY, 2015)

Em síntese, a autora aponta e destaca que os problemas mais comuns de qualidade de dados podem incluir: problemas intrínsecos de qualidade de dados, que dizem respeito ao próprio conteúdo e a estrutura dos dados, como múltiplas fontes para o mesmo dado, credibilidade ou objetividade questionáveis, e má reputação dos dados; problemas de acessibilidade, que podem ser resultado de recursos computacionais precários, falta de controle e segurança do acesso aos dados, má organização da(s) base(s) ou quantidade excessiva de dados; problemas contextuais como dados incompletos ou de baixa relevância e problemas representacionais, ligados ao entendimento e compreensão dos dados pelos usuários. (SIDNEY, 2015)

A autora também pondera que, na medida em que a qualidade de dados vem a ser um dos mais importantes aspectos no âmbito dos esforços de gestão sobre dados e informações de uma empresa, pode também tornar-se um dos mais desafiadores. Em especial para organizações com grandes volumes de dados ou conjuntos de dados muito complexos, pode ser difícil saber por onde começar e como proceder adequadamente em todas as etapas na correção de erros percebidos nos dados. (SIDNEY 2015).

Segundo Burbank (2014), a qualidade de dados só pode ser considerada e, portanto, mensurada ou avaliada, dentro do contexto específico do uso pretendido para os dados. O nível exigido de qualidade de dados para um componente de dados em particular depende especificamente dos processos de negócio que interagem com estes dados. Assim, os mesmos

dados poderiam ser avaliados em diferentes níveis de qualidade, de acordo com as diferentes necessidades, percepções e contextos dos seus diferentes usuários e interlocutores. (BURBANK, 2014, apud SIDNEY, 2015)

De acordo com um estudo realizado nos Estados Unidos (TAYI apud FONSECA et al, 2007), citado por Sidney (2015), a baixa qualidade de dados chegou a representar prejuízos e perdas entre 8-12% dos custos operacionais das empresas avaliadas. Segundo Eckerson (2002), os custos para as organizações referentes a bases de dados sem qualidade ou com erros críticos incluem custos irrecuperáveis, retrabalho de produtos e/ou serviços e perdas de vendas e podem representar entre 10 e 25% do orçamento total da organização. (ECKERSON, 2002, apud SIDNEY, 2015)

No contexto organizacional, a má gestão ou a má qualidade de dados não apenas prejudica a eficiência da tomada de decisão, mas pode gerar também altos custos operacionais com a manutenção e reparo de erros. Santos, 2010, afirma que a boa qualidade da informação é essencial para apoiar o processo decisório em qualquer nível organizacional, e por este motivo é importante que essa informação seja baseada em dados de qualidade. (SANTOS, 2010, apud SIDNEY, 2015)

3.2 Open Data – Dados Governamentais abertos

Tratando-se de dados e informações especificamente no Setor Público, é importante destacar alguns aspectos sobre dados abertos (open data), que, para informações governamentais são entendidos, segundo Ribeiro e Almeida (2008 p.4), como o esforço para “publicação e disseminação das informações do setor público na Web”, permitindo a visualização, reutilização e a integração destes dados por indivíduos de toda a sociedade, de acordo com suas necessidades. (RIBEIRO; ALMEIDA, 2008)

Atualmente, é impossível dissociar o acesso à democracia e a políticas públicas em geral, da transparência na gestão governamental e ao acesso às informações produzidas pelo governo, nos seus diversos níveis de estruturação. A própria sociedade é cada vez mais exigente quanto à

transparência governamental, por isso, o simples acesso às informações já não é o bastante, pois a capacidade de compartilhar, tratar e reutilizar estas informações é fator determinante para acompanhar efetivamente as ações de governo. (RIBEIRO; ALMEIDA, 2008)

Segundo os autores, a efetiva implementação de dados governamentais abertos, dentre outros fundamentos, deverá obedecer a três critérios essenciais: os dados devem ser encontrados e indexados na Web, do contrário eles não existem; os dados devem estar abertos e disponíveis em formato compreensível por máquina, do contrário eles não podem ser reaproveitados; e os dados devem ter sua reuplicação salvaguardada por todos os dispositivos legais, do contrário eles não serão úteis. (RIBEIRO; ALMEIDA, 2008)

Destaca-se que o objetivo de que informações públicas sejam divulgadas como dados abertos é superar as limitações para que as informações do serviço público possam ser facilmente encontradas, acessadas, compreendidas e utilizadas de acordo com os interesses e conveniências diversas dos usuários. A disponibilização de dados governamentais abertos permite a utilização das informações de acordo com a maneira e conveniência do usuário, podendo ser misturadas, combinadas e analisadas para agregar maior valor ao conjunto de dados. (DINIZ, 2010, apud RIBEIRO; ALMEIDA, 2008)

Nesta perspectiva, vários governos com visão de futuro estão demonstrando como o governo pode catalisar o desenvolvimento deste ecossistema através da abertura dos seus próprios conjuntos de dados e da gestão ativa da sua difusão e utilização. Essa é uma tendência a nível mundial. (WORLD ECONOMIC FORUM, 2012)

Pode ser citado, dentre tantos casos recentes, por exemplo, o caso do Governo do Quênia, que lançou o seu novo Portal de Dados Abertos, que inclui dados de públicos detalhados coletados nos últimos 12 anos no país, desde pesquisas sobre o rendimento das famílias, até a localização geográfica de equipamentos públicos, como escolas e unidades de saúde. O portal oferece acesso ilimitado de dados para pesquisadores, desenvolvedores (web e de

software), jornalistas, estudantes, sociedade civil e do público em geral. Organizações civis, desenvolvedores de aplicativos móveis, e grupos de mídia já estão utilizando estes dados para melhorar a compreensão dos padrões populacionais, mapear os serviços públicos e aumentar a transparência governamental em todas as esferas. (WORLD ECONOMIC FORUM, 2012)

3.3 – *Cloud Computing*: Computação em Nuvem

Antes de tratar propriamente do *Big data* em si, é importante apresentar o conceito (e também fenômeno) do *Cloud Computing*, que em tradução literal significa “computação em nuvem”, o que na prática, é tido como o armazenamento digital e remoto de dados, informações, mídias, sistemas ou serviços. As informações são mantidas e processadas na nuvem, ou seja, em um servidor que a disponibiliza para outros usuários remotamente, possuindo algumas importantes facilidades de acesso, processamento e armazenamento de grandes volumes de informações. (CANARY, 2013)

Segundo a Computer Science Corporation – CSC, uma grande multinacional do ramo de tecnologia da computação, a computação em nuvem corresponde à elevada capacidade computacional, flexível e disponível, o que representa uma utilização mais eficiente dos recursos¹. A partir de simples conexões remotas de rede (à nuvem), aspectos como a largura de banda, espaço disponível de armazenamento, poder de processamento, confiabilidade e segurança são garantidos pelo próprio provedor de serviços, facilitando o acesso dos usuários. Bastaria uma conexão à nuvem, e as necessidades, em termos de tecnologias de informação, a nível individual ou corporativo, seriam satisfeitas.

Neste sentido, a computação na nuvem seria capaz de compreender os seguintes recursos:

¹ Disponível em: http://www.csc.com/pt/offerings/63346-o_que_%C3%A9_o_cloud_computing

Quadro 03 – Recursos do Cloud Computing

Recursos Disponíveis	Descrição
<i>Serviço Self-Service</i>	<i>O usuário pode, unilateralmente, requerer ou dispensar capacidades de computação, tais como o tempo do servidor, a capacidade de armazenamento, ou outros, conforme necessário e de forma automática, sem intervenção humana.</i>
<i>Acesso à rede em banda larga</i>	<i>Todas as funcionalidades estão disponíveis através da rede e são acessíveis por meio de mecanismos padrão, que promovem o uso de plataformas-cliente heterogêneas (telefones móveis, laptops, PDAs, smartphones, entre outros).</i>
<i>Pool de recursos</i>	<i>Os recursos de computação são concebidos para servir vários clientes, num modelo multi-tenant, com diferentes recursos físicos e virtuais, distribuídos e alocados dinamicamente. Exemplos de recursos incluem o armazenamento, processamento, memória, largura de banda de rede, máquinas virtuais, entre outros.</i>
<i>Elasticidade</i>	<i>Os recursos podem ser rapidamente alocados e, em alguns casos, de forma automática, para aumentar as capacidades disponíveis ou para disponibiliza-las quando não são utilizadas.</i>
<i>Mensurabilidade</i>	<i>Os sistemas em nuvem devem controlar e otimizar a utilização dos recursos de forma automática, monitorando a sua utilização, de forma adequada ao tipo de serviço, como por exemplo, armazenamento utilizado, processamento efetuado, largura de banda utilizada ou contas de utilizadores ativas. O uso dos recursos deve poder ser monitorizado e controlado de forma transparente, tanto para o fornecedor, como para o usuário do serviço utilizado.</i>

Fonte: Adaptado de CSC, 2016

Em nível corporativo global, essa flexibilidade e elasticidade altera um paradigma do mercado: o que antes necessitava ser adquirido, agora poderia ser alugado. Ao invés de terem que investir em data centers que iriam perdurar durante décadas, e trazer consigo elevados custos, os usuários e as próprias organizações podem simplesmente alugar o poder de processamento, da nuvem, e pagar por ele apenas durante o tempo em que o utilizarem, sob demanda. (CANARY, 2013)

Alguns serviços como o AWS da Amazon ou o Azure da Microsoft já possibilitam que milhares de empresas ao redor do mundo possam armazenar,

processar e analisar quantidades massivas de dados com um custo mais baixo, pagando apenas pelo período em que utilizaram os serviços da nuvem. (WORLD ECONOMIC FORUM, 2012)

É incalculável o volume de dados e informações que a sociedade gera a cada 24 horas. Hoje, estima-se em aproximadamente 3,2 bilhões o número de usuários da internet em todo o mundo. A cada minuto, são feitas mais de quatro milhões de buscas no Google e, neste mesmo intervalo de tempo, mais de 120 horas de vídeo são publicados no YouTube. Somente o Google processa aproximadamente, todos os dias, 20 milhões de gigabytes de dados (Vide Anexo II). O Facebook, que atualmente possui mais de 1,60 bilhões de usuários, gera uma estimativa de 4,75 bilhões de novos conteúdos diariamente entre links, comentários, vídeos, fotos e atualizações de status. Este grande volume de informações vem apenas de três dos inúmeros portais e serviços disponíveis na Internet. Ao imaginarmos a internet como um todo, existe uma profusão enorme de sites, portais de e-commerce, notícias, blogs, serviços, além de dados transacionais e conteúdos não-indexados (a chamada deep web). (CANARY, 2013)

Diante deste contexto, destaca-se um dos principais desafios do *Big Data*: como localizar, armazenar e processar todo esse volume de dados que são produzidos continuamente? E ainda: como armazenar e processar tudo isso de forma economicamente viável? É essencial que se tenha capacidade computacional de armazenamento e processamento (a um custo acessível). (WORLD ECONOMIC FORUM, 2012)

Nesse cenário, o *Cloud Computing* é um viabilizador econômico para trabalho com grandes volumes de informação. Hoje, a nuvem dispõe de um poder computacional virtualmente infinito e ao alcance de qualquer indivíduo ou organização. Ela facilita o acesso a um poder de armazenamento e processamento muito grande, além de flexível, essencial para se falar em termos de *Big Data*. (WORLD ECONOMIC FORUM, 2012)

3.4 Big Data

Tendo em vista o aumento exponencial da quantidade de dados e informações disponíveis, surge e populariza-se rapidamente, a expressão *Big Data*. Segundo o autor, não há ainda uma definição específica e precisa para *Big Data*, entretanto, o termo tem sido frequentemente usado para designar todo um conjunto de tendências tecnológicas que permitam novas abordagens para o tratamento e entendimento de grandes conjuntos de dados e informações para fins de tomada de decisão. (TAURION, 2012)

Segundo o grupo Gartner (2012), o conceito de *Big Data* envolve ativos de alto volume, velocidade e variedade de informação que exigem elevado retorno em termos de custo-benefício e formas inovadoras de processamento de informações para efetivar a tomada de decisão. (GARTNER GROUP, 2012, apud CAMPOS, 2015)

Uma definição bastante ampla é trazida pela IBM, segundo a qual *Big Data* é o termo utilizado para descrever grandes volumes de dados e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas a cada dia. Dentre as muitas razões para que se caminhe em direção a *Big Data*, uma delas é a busca da computação cognitiva, em que empresas como a própria IBM já desenvolvem aplicações de “*machine learning*” que aprendem com dados, e quanto mais dados, mais os algoritmos das máquinas aprendem. (IBM, 2013 apud CAMPOS, 2015).

Outro conceito interessante é trazido pela Internacional Data Corporation (2011) segundo a qual as tecnologias de *Big Data* descrevem uma nova geração de tecnologias e arquiteturas projetadas para extrair economicamente o valor de volumes muito grandes e de uma variedade de dados, o que permitiria tanto a captura, como o tratamento e análise de dados em grande velocidade e volume. (IDC, 2011, apud CAMPOS, 2015).

Quadro 04 – Alguns conceitos e aspectos do *Big Data*

MANYKA, J; et. al. (2011) (McKinsey Global Institute)	“Big Data refere-se a conjuntos de dados cujo tamanho é além da capacidade de ferramentas de software de banco de dados típicos para capturar, armazenar, gerenciar e analisar.”
MCAFEE, A; et. al. (2012) (Harvard Business Review)	“Big Data como uma forma essencial para melhorar a eficiência e a eficácia das organizações de vendas e marketing. Ao colocar Big Data no coração de vendas e marketing, os insights podem ser aproveitados para melhorar a tomada de decisão e inovar no modelo de vendas da empresa, o que pode envolver a utilização de dados para orientar ações em tempo real.”
DEMIRKAN, et. al. (2012) (Decision Support Systems)	“Há o desafio de gerenciar grandes quantidades de dados (Big Data), que está ficando cada vez maior por causa do armazenamento mais barato e evolução dos dados digitais e dispositivos de coleta de informações, como telefones celulares, laptops, e sensores.”
PHELAN, Mike (2012) (Forbes)	“O fenômeno surgiu nos últimos anos devido à enorme quantidade de dados da máquina que está sendo gerado hoje - [...] - juntamente com as informações adicionais obtidas por análise de todas essas informações, que por si só cria outro conjunto de dados enorme.”

Fonte: Canary, 2013 p.22

Ainda, segundo Taurion (2012), *Big Data* é um conjunto de dados que necessariamente possui as seguintes características: Volume, Variedade, Velocidade e Veracidade. Por Volume, entende-se a quantidade de dados, geralmente formados por grandes conjuntos. Variedade tem a ver com os diferentes formatos e estruturas presentes no conjunto de informações. Velocidade representa o quão rápido a informação muda, ou, o quão rápido, novos pontos de informação aparecem, o que impacta o tempo de processamento. Por fim, a Veracidade está relacionada com a qualidade dos dados, ou seja, se eles são corretos e correspondem à realidade ou não.

O autor ainda aponta um quinto “V”, que se refere ao Valor. De uma forma geral, mesmo que os dados tenham volume, variedade, velocidade e veracidade, se não forem dotados de valor intrínseco, o seu processamento, por mais sofisticado que possa ser, não vai trazer resultados para a organização. (TAURION, 2012)

Big Data, então, se trata de um conjunto de tecnologias de informação que são capazes de analisar e processar grandes volumes de dados de diferentes fontes e origens, estruturados ou não estruturados, usando

ferramentas e recursos que sejam capazes de fazer isso em alta velocidade, provendo novos insights e subsídios à tomada de decisões.

Taurion, (2012), ressalta que, atualmente, *Big Data* é visto como um modismo ou *hype*², nas palavras do próprio autor. Entretanto, é um fenômeno ainda recente, e não pode ser ignorado. Quando se estabilizar, segundo o autor, será conceituado apenas como *Data*, com um volume absurdamente grande, mas juntamente com outras características (velocidade de tratamento, variedade de dados, veracidade e valor) que estarão combinados para gerar novos processos. Segundo ele, tratar e analisar dados será tão importante para as organizações, quanto os demais fatores, como recursos humanos, tecnológicos e financeiros, de modo que as organizações simplesmente não viverão mais sem a análise contínua de dados. (TAURION, 2012)

Segundo Pessoa, Freitas e Borges (2016), *Big Data* não pode ser compreendido também sem o desenvolvimento de linguagens visuais adequadas. De acordo com os autores, um site, sistema, ou plataforma digital composto por grandes volumes de dados e por visualizações interativas, complexas ou dinâmicas, pode ser de difícil uso quando a interface e a experiência do usuário não são adequadamente consideradas e projetadas. (PESSOA; FREITAS; BORGES, 2016)

Ao final deste trabalho, consta, em seu Anexo II, material publicitário e informativo, produzido pela equipe do DataViva, em parceria com a “Tinno Comunicação” sobre “Afinal, o que é *Big Data*?” trazendo alguns apontamentos e conceitos interessantes sobre o tema. Segundo os autores,

“[...] A verdade é que o big data não é nem uma ferramenta, nem um conjunto de informações, nem uma técnica ou um serviço: esse termo faz referência a um fenômeno em que o volume de informações disponíveis cresceu muito e de modo muito rápido, tornando necessária a criação de novas plataformas de análise para processar essa grande quantidade de dados. (TINNO COMUNICAÇÃO, 2016, p.2)

² Hype é a promoção exacerbada de uma pessoa, ideia, produto ou serviço. É geralmente definido como o assunto que está "dando o que falar" ou algo sobre o qual todos falam e comentam. Geralmente é algo passageiro, como o assunto da moda. A palavra deriva de hipérbole, figura de linguagem que representa o exagero de algo ou uma estratégia para enfatizar alguma coisa. (<https://pt.wikipedia.org/wiki/Hype>)

3.5 “Ecosystema de dados”

Vimos que, dentro e fora das organizações, os dados e as informações constituem a base essencial para quase todos os processos, rotinas e tomadas de decisão. Por esse motivo, é de vital importância que esses ativos informacionais sejam de qualidade, seguros, e disponíveis, conforme necessário, caso contrário, perdem seu valor. É possível, contudo, agregar valor aos dados através de interações colaborativas.

De acordo com os autores do relatório “*Big Data, Big Impact: New Possibilities for International Development*”, é extremamente relevante a criação de ambientes comuns para a partilha de dados que são gerados diariamente. Segundo os autores, essa colaboração pode favorecer todos os atores envolvidos nesse ambiente de partilha, sendo que existem inúmeros benefícios provenientes da participação colaborativa, mas estes devem estar sempre comprometidos com uma participação transparente, ativa, e aberta. (WORLD ECONOMIC FORUM, 2012)

Este ambiente de partilha, que os autores chamam de *Data Ecosystem* ou “*Ecosystema de dados*”, em tradução literal, envolve, necessariamente, a participação do governo, setor privado e de pesquisa e indivíduos, sendo que possui diferentes diretrizes e\ou incentivos, para cada um destes atores, em diferentes níveis, com critérios e parâmetros bem-estabelecidos, de acordo com o arranjo institucional em questão. (WORLD ECONOMIC FORUM, 2012)

Analisando as diversas possibilidades de dados ou informações e atores presentes em um ecossistema de dados, evidenciam-se também os distintos papéis e incentivos no trabalho com as informações disponíveis no ecossistema. O setor privado, por exemplo, mantém vastos repositórios de dados transacionais, muitos dos quais são “armazéns de dados” ou dados criados como subprodutos de outras transações. Com o uso da telefonia móvel, por exemplo, grande parte destes dados poderia ser associada a indivíduos e

as suas localizações, aprimorando diversos serviços e aplicações. (WORLD ECONOMIC FORUM, 2012)

De forma semelhante, o Poder Público, na maioria dos países, também mantém vastos repositórios, bases e conjuntos de dados, na forma de dados censitários e de recenseamento, indicadores de saúde, informações fiscais e de despesa, entre diversas outras. Entretanto, para transformar os dados em ferramentas de desenvolvimento, uma série de elementos do ecossistema de dados deve estar presente, segundo os autores. Para aqueles indivíduos que geram os dados continuamente, mecanismos devem ser desenvolvidos para garantir a privacidade do usuário e a sua adequada segurança. Ao mesmo tempo, novos modelos de negócios podem ser criados para fornecer incentivos adequados para os atores do setor privado, para compartilhar e utilizar os dados em benefício da sociedade. Alguns destes modelos já existem no ambiente da Internet, por exemplo: dados fornecidos pelos indivíduos e usuários voluntariamente, ou por meio do “*crowdsourcing*”. (WORLD ECONOMIC FORUM, 2012)

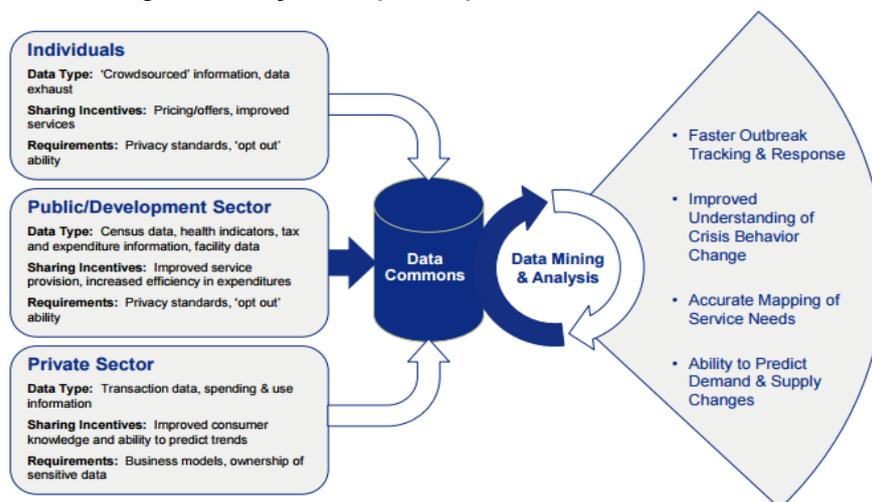
Diversas organizações já se especializam em pesquisas e análises de dados e informações, e a partir disso oferecem novos produtos e serviços, até mesmo sem custos para usuários finais, pois em muitos casos, os dados de tráfego e uso que tais produtos geram, por si só, são valiosos para outros atores do ecossistema. (WORLD ECONOMIC FORUM, 2012)

Segundo os autores, os governos devem tomar a frente e a liderança na definição de políticas, diretrizes e estruturas que protejam a privacidade dos indivíduos e demandem ao setor privado que torne públicos os dados que possui. Organizações ligadas ao desenvolvimento podem apoiar os governos e demonstrar tanto o bem público como o valor comercial que a partilha de dados pode promover. O setor privado, por sua vez, pode criar mecanismos para a partilha de dados que podem beneficiar os indivíduos, de modo cada vez mais rápido e eficiente. (WORLD ECONOMIC FORUM, 2012)

A figura abaixo mostra um modelo do chamado ecossistema de dados, destacando os diferentes atores, incentivos e requisitos, bem como os

resultados esperados da interação colaborativa destes atores. (WORLD ECONOMIC FORUM, 2012)

FIGURA 1 – Dados, incentivos e requisitos dos atores participantes em um modelo de governança composto por um “ecossistema de dados”:



Fonte: World Economic Forum, 2012 pg. 9

Modelos semelhantes ao da figura acima podem postos em aplicação, sobretudo em se tratando de iniciativas de *Big Data*, uma vez que os dados gerados pelo ecossistema poderiam maximizar o impacto da aplicação de recursos públicos (que são escassos), indicando onde tais recursos são mais necessários.

Além disso, os atores de todo o ecossistema têm muito a ganhar com a criação de repositórios de dados abertos (“*Data Commons*”), tendo em vista os demais aspectos já discutidos acerca do *open data*. No entanto, para criação de interfaces diretas junto aos usuários, os autores destacam que o compartilhamento de dados, especialmente quando vinculados a indivíduos, suscita preocupações legítimas, sobretudo no que diz respeito à segurança, privacidade e integridade, que devem ser abordadas e solucionadas para alcançar uma colaboração intersetorial efetiva e frutífera. (WORLD ECONOMIC FORUM, 2012)

Nesta mesma linha, é possível traçar um paralelo entre o modelo de Ecossistema de Dados proposto pelos autores com um conceito já bastante discutido no âmbito da formulação de políticas públicas, sobretudo em

iniciativas colaborativas e contemporâneas de inovação no setor público, o conceito de governança colaborativa. Este, em linhas gerais, pode ser entendido como um arranjo decisório, coletivo, e que envolve tanto atores governamentais quanto atores privados, com o propósito de formular e implementar uma política pública ou gerenciar um programa com fins públicos. (PESSOA, FREITAS, BORGES, 2016) No tópico seguinte, trataremos de aspectos gerais do setor público e adentraremos melhor no conceito de governança colaborativa, e suas implicações.

3.6 Aspectos do Setor Público e da Governança Colaborativa

Segundo Bresser-Pereira, os recursos econômicos e políticos são escassos, mas é possível superar parcialmente tal limitação através do seu uso eficiente pelo Estado na prestação dos serviços públicos. Assim, uma Administração Pública eficiente passa a ter valor estratégico para superar a lacuna que separa as demandas sociais e a capacidade do Governo de satisfazer tais demandas. (BRESSER-PEREIRA, 1998)

A adequação das organizações públicas a se orientarem por seus objetivos e resultados, com a prestação de serviços mais eficientes aos cidadãos, e desempenho devidamente avaliado, foi outra característica marcante em grande parte das reformas administrativas do Estado, que culminaram nos modelos de Administração Pública Gerencial, Nova Administração Pública e Nova Gestão Pública, como denominam alguns autores. (GUERRA, 2009)

Além disso, a nova forma de Administração tomava emprestados alguns dos mecanismos gerenciais e avanços que perpassaram as organizações privadas durante todo o século XX, sem, contudo, deixar de visar a satisfação do interesse público. Pode-se dizer que as organizações públicas vêm se modernizando e absorvendo práticas e modelos de gestão oriundos das organizações privadas, no intuito de promover aumento e maior qualidade de resultados, trazendo para si princípios permeados pela eficiência e eficácia, além de critérios de técnica e racionalidade, já consagrados na esfera privada. (GUERRA, 2009)

Segundo Bresser-Pereira:

“À nova administração pública não basta ser efetiva em evitar o nepotismo e a corrupção: ela tem de ser eficiente ao prover bens públicos e semi-públicos que cabe ao Estado diretamente produzir ou, indiretamente, financiar [...] a partir da crítica ao modelo burocrático, no qual haveria um excessivo apego às regras e procedimentos, a nova administração pública defende a flexibilização dos meios e a orientação da organização e dos agentes públicos para o alcance de resultados.”
(BRESSER – PEREIRA, 1998, p.27)

Assim, na perspectiva de uma gestão pública pautada por princípios ligados à eficiência, percebe-se cada vez mais a expansão e consolidação do uso das tecnologias de informação e comunicação (TIC's) no âmbito da Administração Pública, como instrumento para aumento da produtividade e redução de custos. O uso adequado das TIC's poderia aprimorar a sistematização e compreensão dos dados e informações que a organização possui e pode permitir à gestão pública a automatização e aprimoramento dos seus fluxos de informações internos e externos, fornecendo-lhe mais recursos para tomar decisões mais acertadas e desenvolver estratégias mais alinhadas com a realidade. No setor público, gerir a informação torna-se ainda mais importante, pois pode culminar na eficiência e efetividade dos serviços prestados à sociedade. (SIDNEY, 2015).

Seguindo na mesma linha de alguns destes apontamentos, parte dos instrumentos de planejamento governamental, como o Plano Mineiro de Desenvolvimento Integrado (PMDI 2011-2030), apresentava, à época, um destaque à indução de uma forte agenda de inovação como estratégia para o desenvolvimento econômico e social. Ademais, neste documento, é também dado grande enfoque e atenção ao aspecto tecnológico dos tempos atuais e os seus impactos na gestão governamental, dada sua forte associação às mudanças recentes nos padrões econômicos, e a necessidade de respostas, por parte do poder público, a estes novos desafios. (MINAS GERAIS, 2011, apud PESSOA; FREITAS; BORGES, 2016)

Neste contexto e também ilustrando algumas das tendências abordadas neste tópico, Rezende (2014), aponta que entre os anos de 2002 a 2014, o Governo do Estado de Minas Gerais esteve envolvido em “um contexto de

implementação da Nova Administração Pública, propugnando um modelo de gestão orientada para resultados” (p. 5), e que, além de enfatizar princípios ligados à eficiência, envolve diversos atores não-governamentais, incluindo ainda iniciativas como:

- (I) *a descentralização dos controles gerenciais (a fim de prover uma maior autonomia dos órgãos e entidades);*
- (II) *a responsabilização e flexibilização de procedimentos e processos;*
- (III) *a introdução de mecanismos de mercado na gestão pública;*
- (IV) *o fortalecimento do uso de novas tecnologias;*
- (V) *a distinção mais específica entre formulação e implementação de políticas públicas;*
- (VI) *uma maior transparência governamental (por exemplo, na realização dos gastos públicos).* (PESSOA; FREITAS; BORGES, 2016 p. 5)

No que tange especificamente à formulação de políticas públicas, contudo, destaca-se a necessidade de uma definição clara dos objetivos, das metas a serem alcançadas e dos programas e ações necessários, bem como as linhas de atuação a serem aplicadas para atingir os resultados esperados, para toda e qualquer ação governamental eficiente e eficaz. A definição clara dos objetivos não apenas representa uma orientação por resultados, mas também ajuda na elaboração das linhas de atuação a serem aplicadas. (GUERRA, 2009)

Nesta mesma linha, segundo o SEBRAE, 2014, em material informativo aos gestores municipais, acerca da formulação de políticas públicas, a primeira providência a ser tomada quando uma situação é vista como problema, – e, por isso, é incluída na agenda governamental –, em um processo de formulação de políticas públicas, é definir as linhas de ação e os recursos que serão utilizados para resolver a questão. Os autores destacam, no entanto, que existem, e podem ocorrer, embates políticos entre grupos que percebem tais linhas de ação como sendo favoráveis ou contrárias a seus interesses. (SEBRAE, 2014)

Assim, inicialmente, devem ser definidos os objetivos da política pública, quais serão os programas desenvolvidos e as metas a serem alcançadas. Mesmo assim, ao final do processo de definição destes três itens, várias propostas de ação poderão ser rejeitadas, pois tais escolhas, além de ter repercussões junto as demais esferas de poder e aos grupos sociais, devem

levar em conta as considerações do corpo técnico e institucional da administração pública, inclusive no que se refere aos recursos – materiais, econômicos, técnicos, pessoais, entre outros. (SEBRAE, 2014)

Segundo os autores, um bom processo de elaboração de políticas públicas segue, em geral, os seguintes passos:

- Conversão de estatísticas em informação relevante para o problema;
- Análise das preferências dos atores; (e suas possíveis interações)
- Ação baseada no conhecimento adquirido. (SEBRAE, 2014)

Para a elaboração de propostas, sugere-se, por exemplo, que o responsável pela preparação da política pública deve ser reunir com os atores envolvidos no contexto no qual ela será implantada e demandar a estes atores que, juntos, apontem as melhores formas de se proceder. Também é interessante (e recomendável) a definição de alternativas, caso a proposta apontada inicialmente se mostre inviável. Tais procedimentos proporcionam à autoridade gestora uma série de informações que podem servir como base para apontar os caminhos desejados pelos segmentos sociais, auxiliando no processo de tomada de decisão e contribuindo com a legitimidade das propostas apresentadas, junto a seu público-alvo. (SEBRAE, 2014)

O que se observa, em todo caso, é que os papéis e a interação entre todos os diferentes atores é um ponto crucial. Alguns estudos e desenvolvimentos teóricos recentes têm buscado discutir os conceitos e fatores que condicionam e impactam os resultados dos processos de interação entre diferentes atores, públicos e privados, de maneira sistematizada, normalmente a partir de estudos de caso de políticas específicas. Nesse aspecto, o conjunto de atores que pode interagir conjuntamente na formulação e na implementação de uma política pública (ou mesmo de um projeto de inovação no setor público), pode envolver atores variados, governamentais e não governamentais, com objetivos de se solucionar situações complexas. (PESSOA, FREITAS, BORGES, 2016)

Ainda, de acordo com Pessoa, Freitas e Borges (2016), modelos teóricos de formulação e implementação de políticas públicas sempre abordaram a questão da interação entre os diferentes atores envolvidos nestes

processos. Segundo os autores essa interação também é estudada à luz dos recentes modelos de gestão pública, emergindo assim, neste campo de estudos, o conceito de governança colaborativa. Os autores ressaltam, contudo, tratar-se de um conceito controverso e não-consensual de acordo com a literatura recente em políticas públicas. (PESSOA, FREITAS, BORGES, 2016)

Ansell & Gash, 2008, definem a governança colaborativa como:

“Arranjo administrativo governamental no qual uma ou mais agências públicas engajam atores não-estatais diretamente em um processo de tomada de decisão coletiva que é formal, orientado para o consenso e deliberativo, visando elaborar ou implementar políticas públicas, ou gerenciar programas ou ativos públicos” (ANSELL, GASH, 2008, apud PESSOA, FREITAS, BORGES, 2016 pg.5)

Citando ainda Ansell e Gash, 2008, os autores supracitados destacam a abordagem da governança colaborativa como oportuna à implementação de projetos inovadores no setor público, sendo esta abordagem um enorme avanço em relação ao modelo gerencialista simples e aos modelos de interação baseados em conflito, usualmente adotados em estudos sobre a formulação e implementação de políticas públicas (p.5). Segundo os autores, este seria um conceito bastante abrangente, que busca observar aspectos importantes de diversas espécies de processos de interação colaborativos e os seus respectivos resultados, envolvendo o papel e a atuação do Poder Público. (ANSELL, GASH, 2008, apud PESSOA, FREITAS, BORGES, 2016)

Emerson, et al, 2012, por sua vez, definem o conceito de governança colaborativa, de modo mais amplo, da seguinte forma:

“Processos e estruturas de gestão e tomada de decisão em políticas públicas que engajam pessoas de maneira construtiva através das fronteiras das agências públicas, níveis de governo e/ou esferas públicas, privadas e civis, no sentido de alcançar um propósito público que não poderia ser alcançado de outra forma.” (p. 2) – tradução dos autores. (EMERSON et al. 2012, apud PESSOA, FREITAS, BORGES, 2016, pg. 5)

Como veremos adiante, a partir da apresentação e do presente estudo de caso da plataforma DataViva, este projeto contou com o envolvimento e a participação de diversos representantes do sistema econômico mineiro,

incluindo centros e agências de fomento a pesquisas, poder público, setor produtivo, empresas e universidades (PESSOA; FREITAS; BORGES, 2016).

Iniciado em 2011, pelo Governo de Minas Gerais, o projeto de desenvolvimento da Plataforma DataViva constitui a elaboração uma ferramenta tecnológica de suporte à tomada de decisão, totalmente aberta e desenvolvida em software livre para o acesso de todos. O seu objetivo seria permitir a identificação dos processos de diversificação e complexidade produtiva e econômica do Estado de Minas Gerais através da análise de bases de dados públicos e oficiais brasileiros, utilizando aportes tecnológicos e inovações metodológicas no tratamento e apresentação da informação. (PESSOA; FREITAS; BORGES, 2016)

Nesta perspectiva, é importante observar a interação entre estes atores e a governança presente no projeto, sob a perspectiva de governança colaborativa, e constituindo-se também como um modelo similar ao ecossistema de dados, apontado no relatório do *World Economic Forum* como fator essencial e oportuno ao desenvolvimento de aplicações úteis do *Big Data*, principalmente envolvendo o setor público e os governos.

4 METODOLOGIA

Segundo Marconi e Lakatos, 2010, métodos são o conjunto das atividades racionais e sistemáticas que permitem alcançar objetivos, traçando os caminhos a ser seguidos, detectando erros e auxiliando as decisões com maior segurança e economia. Estabeleceremos aqui critérios atinentes à metodologia da pesquisa para o estudo ora proposto.

O trabalho se desenvolve mesclando características de uma pesquisa qualitativa descritiva e bibliográfica, com um estudo de caso. Para tanto, a coleta de informações e conceitos foi realizada por meio de revisão bibliográfica com levantamento de publicações relacionadas ao objeto de estudo.

Sob essa perspectiva, o estudo busca apresentar os conceitos e opiniões de diversos autores acerca do que vem a ser *Big Data*, caracterizando este conceito. Além disso, delimita, em seu referencial teórico, alguns conceitos relativos à: Dados e Informações de modo amplo, Qualidade de dados, Tomada de Decisão, Dados abertos (*Open Data*) e Governança Colaborativa – ecossistema de dados, que se relacionam ao tema proposto e propiciam suporte às considerações acerca de sua aplicabilidade e relevância, no caso concreto em análise.

Assim, a partir da delimitação do referencial teórico, é apresentado *estudo de caso* de *Big Data* no Setor Público, com o caso da Plataforma DataVIVA. A partir desse caso prático, buscou-se verificar os aspectos de um Projeto de Big Data de iniciativa pública, tal como se deu, no âmbito da estrutura governamental, evidenciando os aspectos ligados a governança colaborativa, exemplificando e demonstrando uma potencial aplicação da Gestão da Informação, buscando ponderar sobre as oportunidades e potencialidades destas iniciativas, bem como os desafios e entraves à sua implementação, à luz do referido estudo de caso.

Por fim, serão apresentadas conclusões relacionadas aos principais aspectos observados dessas aplicações. A partir disso, será possível elencar

percepções oriundas do estudo, ressaltando o seu objetivo maior, que é avaliar a aplicabilidade de determinadas soluções informacionais (*Big Data*) a uma realidade específica, em âmbito organizacional, do setor público.

5 ESTUDO DE CASO: Plataforma DataVIVA

O DataViva é uma plataforma de pesquisa e visualização de dados e informações socioeconômicas³, criada pelo Governo de Minas Gerais, em parceria com pesquisadores do Instituto de Tecnologia de Massachusetts – MIT, com a colaboração de diversos órgãos e entidades públicas e privadas. Desenvolvida em software livre, aberto ao acesso de todos, e com informações em português e inglês, a plataforma foi criada com o objetivo de contribuir e fornecer suporte informacional para a implementação de políticas públicas, investimentos públicos e privados, bem como para a realização de pesquisas aplicadas e estudos acadêmicos. (PESSOA; FREITAS; BORGES, 2016)

Lançada em sua primeira versão no fim do ano de 2013, a plataforma possibilita a visualização interativa e dinâmica de bases de dados públicos, com séries históricas dos últimos anos, sobre a economia, comércio exterior, indústrias, localidades, mercado de trabalho, educação e ocupações disponíveis, dentre outras, para todos os municípios brasileiros, podendo ser analisadas em diferentes parâmetros de regionalização. Através de diferentes aplicações, é possível obter mais de 1 bilhão de visualizações distintas para as informações, de modo que a plataforma ofereça aos seus usuários uma experiência única de acesso a um robusto conjunto de bases de dados públicos oficiais brasileiros. (PESSOA; FREITAS; BORGES, 2016)

Figura 2: Página Inicial da Plataforma DataViva – “A maior plataforma de visualização de dados sociais e econômicos do Brasil. Totalmente aberta e livre”



Fonte: DataViva

³ Disponível em: <http://dataviva.info/pt/>

Segundo a página da plataforma, um dos pilares de concepção do DataViva é a tecnologia do *Big Data*, sob a seguinte definição: “técnica de geração de conhecimento e inteligência a partir do processamento de um grande volume de dados”.

Figura 3: Exemplo de dados disponíveis no DataViva sobre o comércio exterior – Ranking de Produtos Exportados, por valor (em USD \$) (BRASIL, 2015)

Rankings

Localidade ▾ Ocupação Atividades Econômicas Produto Parceiro Comercial Universidade Ensino Superior Curso Básico

Seção Posição 2015 Pesquisar:

Ano	ID HS	Produto	Complexidade do Produto	Exportações	Importações	Crescimento Nominal das Exportações (1 ano)	Crescimento Nominal das Exportações (5 anos)
2015	021201	Soja	-1,748	USD 21 Bilhões	USD 109 Milhões	-10% 📉	13% 📈
2015	052601	Minério de Ferro	-1,571	USD 14,1 Bilhões	USD 8,15 Mil	-45% 📉	1% 📈
2015	052709	Petróleo Cru	-2,033	USD 11,8 Bilhões	USD 7,38 Bilhões	-28% 📉	5% 📈
2015	041701	Açúcar in Natura	-2,004	USD 7,64 Bilhões	USD 1,74 Milhão	-19% 📉	-2% 📉
2015	010207	Carne de Aves	0,042	USD 6,38 Bilhões	USD 9,66 Milhões	-10% 📉	5% 📈
2015	042304	Farelo de Soja	-1,38	USD 5,82 Bilhões	USD 1,46 Milhão	-17% 📉	5% 📈
2015	020901	Café	-1,867	USD 5,57 Bilhões	USD 67,1 Milhões	-8% 📉	8% 📈
2015	104703	Pastas Químicas de Madeira À Soda...	-0,392	USD 5,34 Bilhões	USD 307 Milhões	9% 📈	12% 📈
2015	021005	Milho	-1,532	USD 5,01 Bilhões	USD 44,6 Milhões	27% 📈	31% 📈
2015	178802	Aviões	-0,711	USD 4,09 Bilhões	USD 1,09 Bilhão	19% 📈	1% 📈

Fonte: DataViva

Planejado inicialmente para auxiliar a política de desenvolvimento econômico do governo do Estado de Minas Gerais, o projeto da Plataforma surgiu a partir da necessidade de tornar o acesso às bases de dados públicas brasileiras mais fácil e intuitivo, e hoje possui diversas aplicações, para diferentes públicos, que vão desde agentes governamentais ligados ao desenvolvimento, quanto estudantes ou empreendedores, buscando novas oportunidades de negócio. (PESSOA; FREITAS; BORGES, 2016)

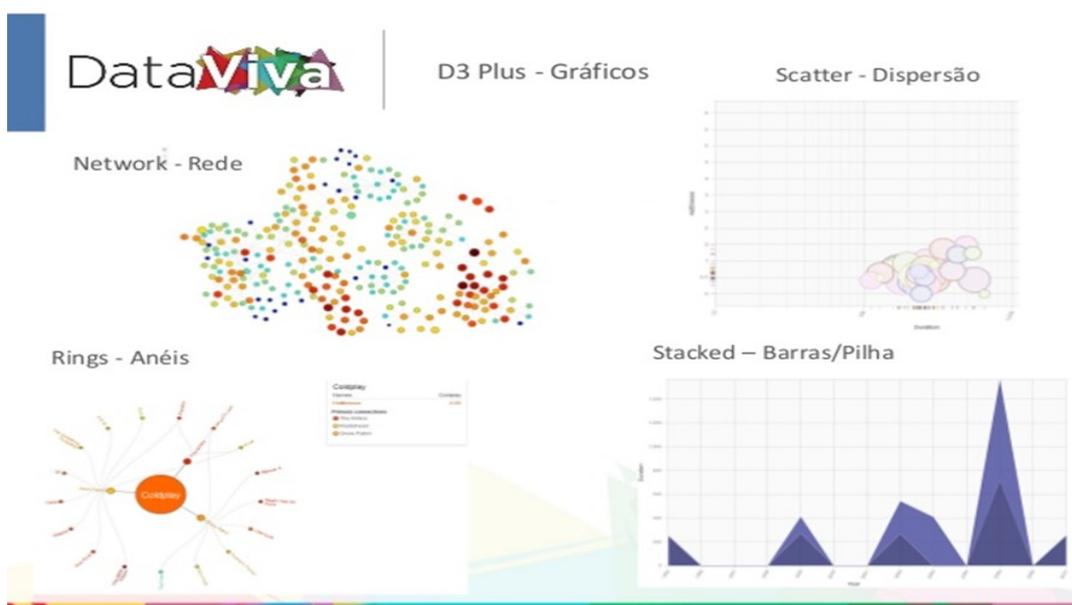
Em sua primeira versão *beta*, lançada em 2013, a plataforma consistia em oito diferentes aplicativos, cada um dos quais permitindo diferentes formas de visualizar os conjuntos de dados disponíveis. Algumas das referidas aplicações são descritivas, isto é, mostram dados agregados em vários níveis, de forma simples e comparativa. Outras são prescritivas, usando cálculos que

sulfatos químicos. Este é um excelente panorama para a visualização, o estudo e o acompanhamento de atividades e políticas econômicas, a curto, médio e longo prazos.

O software usado no DataViva tem código aberto e está disponível para acesso no site *GitHub*⁴. O sistema foi construído utilizando a tecnologia *Python*, com o *framework* de desenvolvimento *Flask*. Todo o processamento de dados e cálculos são feitos utilizando o *Python*, com bibliotecas de programação codificadas nesta mesma linguagem. Outros governos, institutos de pesquisa e empresas podem adaptar o software à visualização de diferentes bases de dados, se assim o quiserem. (PESSOA; FREITAS; BORGES, 2016)

A visualização dos gráficos também é feita por meio de HTML5 e CSS3, baseado em um *framework* chamado D3, o qual, utilizando *Javascript*, cria uma interação com o usuário durante sua navegação, por meio da interface *web*. Utilizando-se dos diversos filtros e aplicativos disponíveis na plataforma é possível criar milhões de visualizações, por meio da combinação de dimensões como localidade, atividade econômica, ocupação, entre outros. (PESSOA; FREITAS; BORGES, 2016)

Figura 5 – Alguns tipos de visualização de dados e informações disponíveis na plataforma.



Fonte: DataViva

⁴ <https://github.com/DataViva/>

5.1 Histórico e objetivos do DataViva

O DataViva começou a ser desenvolvido em 2011. O site foi aberto em setembro de 2013, e inaugurado oficialmente em dezembro do mesmo ano. Hoje, a plataforma ainda está oficialmente em fase de testes, sendo possível ver a palavra “beta” ao lado do logo, ao acessar a plataforma, pois a perspectiva é que ainda sejam agregadas diversas novas bases de dados e informações no futuro.

Apenas para a sua implementação, o projeto custou cerca de 1,2 milhão de reais, à época, financiados pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig). As primeiras bases de dados a compor a plataforma foram disponibilizadas pelos ministérios do Trabalho e Previdência Social (MTPS), do Desenvolvimento, Indústria e Comércio Exterior (MDIC) e da Educação. Em sua primeira versão, o DataViva apresentava dados relativos à exportação de 1.256 produtos (da Secretaria de Comércio Exterior do Ministério do Desenvolvimento, Indústria e Comércio Exterior – SECEX/MDIC) e relativos a 614 ocupações em 427 atividades econômicas (da Relação Anual de Informações Sociais do Ministério do Trabalho e Previdência Social – RAIS/MTPS). (PESSOA, FREITAS, BORGES; 2016)

Na sua segunda versão, foram incluídos dados relativos à importação de produtos (cobrindo assim todo comércio internacional brasileiro) e também dados relativos a cursos básicos, técnicos e superiores (Ministério de Educação). Atualmente, os dados de exportação estão disponíveis para os anos de 2000 a 2016, os dados de ocupação para os anos de 2002 a 2014, para todos os 5.567 municípios brasileiros e os dados de educação para os anos de 2007 a 2015. São mais de 4 milhões de registros de comércio internacional por ano, mais de 50 milhões de registros de empregados por ano e mais de 40 milhões de registros de matrículas. (PESSOA, FREITAS, BORGES; 2016)

Possibilitar que todo este volume de dados e informações, inicialmente desconexas, se transforme em informações e *insights* relevantes é

o verdadeiro desafio de uma plataforma elaborada sob o conceito de big data para agregar valor e extrair significado de bases de dados públicas. Segundo Pessoa, Freitas e Borges, 2016, de modo muito mais amplo, o projeto de construção da plataforma DataViva teve os seguintes objetivos:

- a) *Incorporar, na cultura de gestão estratégica nacional, conhecimentos necessários ao uso de uma plataforma aberta constituída de dados, software, métodos e processos para o estudo das cadeias de valor presentes na economia, de seu inter-relacionamento e das tecnologias a elas subjacentes;*
- b) *Destacar a agenda de inovação do estado (de Minas Gerais), focando nas instituições de pesquisa, universidades e agentes públicos, considerando a oportunidade de diversificar e aumentar a complexidade da estrutura produtiva e habilidades necessárias para gerar um desenvolvimento econômico sustentável;*
- c) *Implantar, em órgãos de planejamento estratégico, métodos e técnicas de análise das cadeias de valor presentes na economia, de seu inter-relacionamento, das tecnologias a elas subjacentes e das possibilidades de inovação oferecidas pelo contexto, no mercado mundial;*
- d) *A partir das técnicas e métodos citados no item anterior, identificar e analisar as estruturas produtivas e suas interações nos municípios, com a finalidade de manter um crescimento para a economia de forma sólida, com a geração de emprego de qualidade e redução das desigualdades socioeconômicas no país;*
- e) *Adquirir independência e competência para uso do DataViva como suporte a essas análises, formar multiplicadores e desenvolver as condições necessárias para propor, com evidências mensuráveis, prioridades estratégicas e caminhos para o adensamento e aumento do valor adicionado nessas cadeias dentro do estado;*
- f) *Estabelecer mecanismos em rede para a arregimentação de talentos criativos, dentro e fora do governo, visando ao estabelecimento de um capital humano geograficamente distribuído. (PESSOA; FREITAS; BORGES, 2016, p. 7)*

A execução do projeto de implantação do DataViva teve uma forte governança colaborativa em seu desenvolvimento e implantação, contando com o apoio de vários órgãos e entidades, além das instituições que compõem o sistema de desenvolvimento econômico do estado, juntamente com a Fundação João Pinheiro, e diversos outros atores do meio acadêmico e empresarial. Aspectos mais detalhados da governança envolvida na concepção

do projeto serão vistos adiante em item específico, assim que apresentadas as bases de dados e de informações que dão suporte à plataforma.

5.2 Bases de dados/informações componentes

Conforme visto até aqui, o DataViva objetiva disponibilizar conhecimentos altamente específicos da economia brasileira, usando, para isso, dados oficiais públicos, em séries históricas atualizadas, de todo o país. As bases de dados públicos oficiais que compõem a estrutura da plataforma DataViva são as seguintes (as informações abaixo constam também na página do DataViva):

A. Relação Anual de Informações Sociais (RAIS) –

Fonte: Ministério do Trabalho e Emprego

A RAIS⁵ é um registro administrativo com periodicidade anual criado com a finalidade de suprir as necessidades de controle, estatística e informação às entidades governamentais da área social. Constitui um instrumento imprescindível para a caracterização e para o acompanhamento do mercado de trabalho formal e o cumprimento das normas legais.

Todo o cenário profissional do país é retratado a partir de informações declaradas pelas próprias empresas sobre sua situação em 31 de dezembro do ano anterior, informando dados sobre seus empregados, atividade econômica desenvolvida, entre outras características. Atualmente, segundo o IBGE, a RAIS cobre aproximadamente 97% do universo do mercado formal brasileiro.

A partir dos dados da RAIS, o DataViva apresenta informações sobre o número de empregos, salários e número de estabelecimentos. O nível de agregação, em termos geográficos, pode ser feito por unidades da federação, mesorregiões, microrregiões e municípios; em termos de setores, seção, divisão e classes.

B. Banco de Dados da Secretaria de Comércio Exterior - SECEX/MDIC

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior – MDIC

⁵ Acesso em: <ftp://ftp.mtps.gov.br/pdet/microdados>

O Banco de Dados da SECEX/MDIC⁶ possibilita a análise histórica do comportamento do intercâmbio comercial brasileiro, um dos mais importantes indicadores de desempenho da economia. Retrata e avalia o movimento comercial do país com as demais nações do mundo, englobando as relações e produtos movimentados no âmbito do comércio exterior brasileiro.

Constitui-se também como um instrumento para tomada de decisão e determinação de diretrizes econômicas por parte do Governo e permite aos agentes envolvidos na atividade o melhor planejamento das suas ações pela análise dos dados concretos das exportações e importações, aumentando as oportunidades de desenvolvimento do comércio externo.

A partir dos dados do SECEX/MDIC, o DataViva apresenta informações das exportações e importações brasileiras, de 2000 em diante com periodicidade mensal. As principais variáveis apresentadas são: valor das exportações e das importações por município, unidade da federação produtora, municípios exportadores e país de destino ou origem. Os dados são divulgados para os seguintes recortes geográficos: Brasil, unidades da federação, mesorregiões, microrregiões e municípios. Por produto, através de códigos específicos, de acordo com o Harmonized System (HS Code), agregados em: seção, capítulo e posição.

C. **CENSO ESCOLAR** - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP

Fonte: Ministério da Educação – MEC

O Censo Escolar é um levantamento de dados estatístico-educacionais⁷ de âmbito nacional realizado todos os anos e coordenado pelo INEP. Ele é feito com a colaboração das secretarias estaduais e municipais de Educação e com a participação de todas as escolas públicas e privadas do país.

Trata-se do principal instrumento de coleta de informações da educação básica, que abrange as suas diferentes etapas e modalidades:

⁶ Acesso em: <http://www.mdic.gov.br//sio/interna/interna.php?area=5&menu=5319>

⁷ Acesso em: <http://portal.inep.gov.br/basica-levantamentos-acessar>

ensino regular (educação Infantil e ensinos fundamental e médio), educação especial e educação de jovens e adultos (EJA). O Censo Escolar coleta dados sobre estabelecimentos, matrículas, funções docentes, movimento e rendimento escolar.

A partir do Censo Escolar, o DataViva apresenta informações sobre o número de matrículas em cada ano/série do ensino básico, agregados por município, microrregião, mesorregião ou unidade da federação. Dados sobre cursos profissionalizantes também pode ser visualizados, neste caso especificando o curso nos mesmos níveis geográficos da educação básica. As escolas também estão divididas por tipo de dependência administrativa e os cursos podem ser filtrados por turno.

D. **CENSO DO ENSINO SUPERIOR** - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP

Fonte: Ministério da Educação – MEC

Anualmente, o Inep também realiza a coleta de dados ⁸ sobre a educação superior com o objetivo de oferecer à comunidade acadêmica e à sociedade em geral informações detalhadas sobre a situação e as grandes tendências do setor.

O Censo da Educação Superior reúne informações sobre as instituições de ensino superior, seus cursos de graduação presencial ou à distância, cursos sequenciais, vagas oferecidas, inscrições, matrículas, ingressantes e concluintes, além de informações sobre docentes, nas diferentes formas de organização acadêmica e categoria administrativa.

A partir dos dados do Censo do Ensino Superior, o DataViva apresenta informações sobre estudantes (ingressantes, matrículas e concluintes) para todos os cursos de nível superior, organizadas por localidades ou por universidades. Os níveis de agregação para as localidades são município, microrregião, mesorregião e unidade da federação. As

⁸ Acesso em: <http://portal.inep.gov.br/basica-levantamentos-acessar>

universidades (e demais organizações acadêmicas) também estão divididas por tipo de dependência administrativa.

E. Banco de Dados Estatísticos de Comércio de Produtos das Nações Unidas -UN Comtrade

Fonte: Divisão de Estatísticas das Nações Unidas

O Banco de Dados Estatísticos de Comércio de Commodities das Nações Unidas⁹ (UN Comtrade) contém estatísticas detalhadas sobre importações e exportações relatadas pelas autoridades estatísticas de cerca de 200 países ou áreas. O UN Comtrade é considerado o banco de dados mais abrangente disponível de comércio internacional, com mais de 1 bilhão de registros. O banco de dados é constantemente atualizado: sempre que os dados de comércio são recebidos de autoridades nacionais, são padronizados pela Divisão de Estatísticas das Nações Unidas e então adicionados ao UN Comtrade.

No DataViva, as informações do UN Comtrade são: origem e destino das exportações e valor exportado detalhado por produto. Ressalta-se que no entanto, os dados não estão disponíveis para visualização direta, sendo utilizados somente para o cálculo de alguns indicadores no DataViva.

A partir de todas estas bases de dados, é possível extrair e cruzar diferentes informações, antes desconexas ou desagregadas. Sem dúvidas a Plataforma é capaz de agregar um enorme valor aos conjuntos de dados que a integram, ao tornar possível uma vasta e complexa miríade de visualizações para estes dados. Além disso, Davenport (1998) já apontava que melhorar a apresentação da informação é uma das chaves para lhe agregar valor. Uma apresentação que cause impacto positivo faz com que a informação seja respeitada, ao passo que uma apresentação pobre ou pouco atraente só pode causar rejeição. (DAVENPORT apud ASSIS, 2006)

Segundo Pessoa, Freitas e Borges, 2016, a próxima fase de evolução do DataViva contempla a incorporação novas bases de dados, incluindo dados

⁹ Acesso em: <http://comtrade.un.org/db/default.aspx>

de saúde (DataSUS) e já se planeja, ainda em fase de negociação, incorporar bases de dados de compras públicas, disponíveis nos órgãos de controle; dados de ciência, tecnologia e inovação; e dados de movimentação de toda a produção nacional, conforme registros das Notas Fiscais Eletrônicas – NF-E’s das Secretarias de Fazenda e Tesouro.

5.3 Governança Colaborativa do Projeto

O desenvolvimento do DataViva contou com o envolvimento, colaboração e troca de informações e apoio técnico de atores diversos, dentro e fora do Governo do Estado de Minas Gerais. É interessante observar tais interações sob as perspectivas de uma Governança Colaborativa e também de um Ecossistema de Dados, pois este arranjo colaborativo terá se mostrado essencial ao êxito da implantação do projeto da plataforma. (PESSOA; FREITAS; BORGES, 2016)

O projeto se deu a partir de uma parceria do Governo de Minas Gerais com uma intensa troca de informações (científicas e tecnológicas) e apoio técnico com o MIT Media Lab, do *Massachusetts Institute of Technology*, bem como uma colaboração interfederativa com os ministérios do Desenvolvimento, Indústria e Comércio Exterior (MDIC), Trabalho e Previdência Social (MTPS) e Educação (MEC) para obtenção das bases de dados. A programação e o desenvolvimento da plataforma, por sua vez, foi desenvolvida por meio de um contrato de prestação de serviços com a *Datawheel*, *spin-off* do MIT Media Lab e detentora das patentes e tecnologias de visualização de dados utilizadas pela plataforma. (PESSOA; FREITAS; BORGES, 2016)

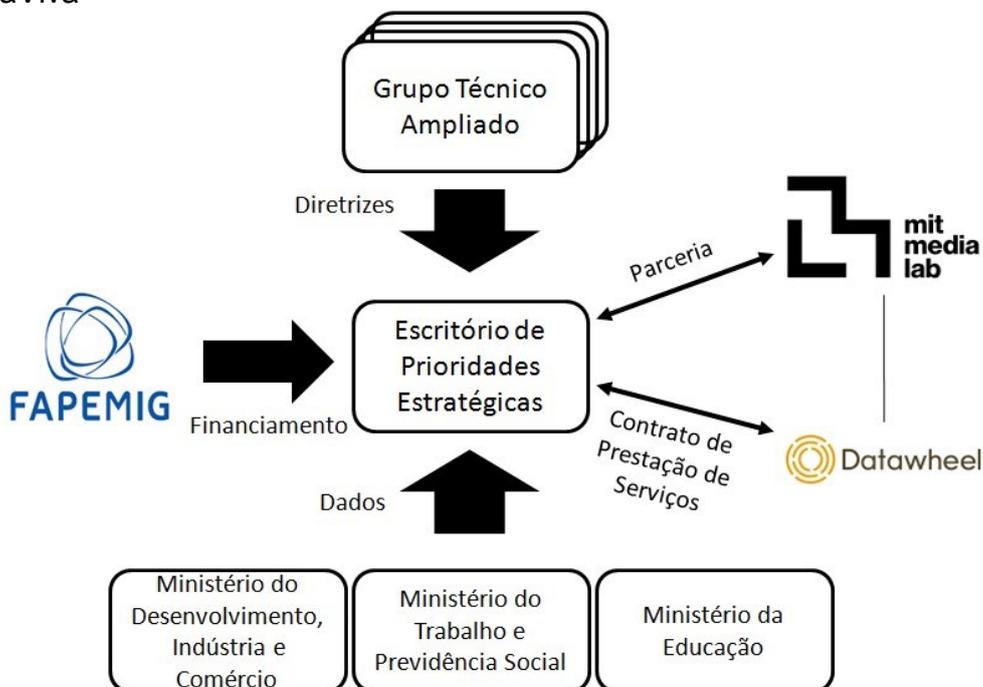
Inicialmente, o responsável pela condução do projeto a nível governamental, foi o Escritório de Prioridades Estratégicas Governo do Estado de Minas, suportado financeiramente por um financiamento direto da agência de fomento à pesquisa e inovação do estado, a Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Cumpre destacar, inclusive, que após a extinção do Escritório, em 2015, a FAPEMIG absorveu toda a

execução e manutenção do projeto bem como a equipe do DataViva, a partir deste mesmo ano. (PESSOA; FREITAS; BORGES, 2016)

Por sua vez, todas as principais diretrizes do projeto, à época de seu início, a partir do ano de 2011, técnicas e/ou institucionais, eram deliberadas, discutidas e estabelecidas a partir do diálogo, das trocas de informações e das interações do chamado “Grupo Técnico Ampliado” (GTA), que foi composto à época por inúmeras entidades governamentais e não-governamentais, e constituiu-se como, portanto, a instância primária de deliberação, colaboração mútua e participação em todas as principais decisões referentes ao conteúdo, aos objetivos e aos rumos do projeto, para além do Escritório de Prioridades Estratégicas, que encabeçava essa iniciativa no Governo. (PESSOA; FREITAS; BORGES, 2016)

A figura 6 abaixo, elaborada por Pessoa, Freitas e Borges (2016), ilustra o Grupo Técnico Ampliado, bem como a gama de relações envolvidas na interação dos diferentes atores e a relação entre as entidades envolvidas na elaboração do projeto DataViva.

Figura 6: Relação entre as entidades envolvidas na concepção do projeto DataViva



Fonte: PESSOA; FREITAS; BORGES, 2016, p.16

De acordo ainda com os autores, o Grupo Técnico Ampliado reunia, ao todo, a participação de 24 organizações e entidades, públicas e privadas, relacionadas de alguma forma, direta ou indiretamente, à política de desenvolvimento econômico do estado de Minas Gerais.

Participaram do Grupo Técnico Ampliado de desenvolvimento da plataforma DataViva os seguintes atores:

- a) Órgãos de governo, tais como: Secretaria de Estado de Desenvolvimento Econômico; Secretaria de Estado da Fazenda; Secretaria de Estado de Desenvolvimento Regional e Urbano; Secretaria de Estado de Educação; Secretaria de Estado de Ciência, Tecnologia e Ensino Superior; Secretaria de Estado de Trabalho e Emprego; Secretaria de Estado de Planejamento e Gestão; Secretaria de Estado de Transportes e Obras Públicas; Instituto de Desenvolvimento do Norte e Nordeste de Minas Gerais; Instituto de Desenvolvimento Integrado de Minas Gerais, para citar os principais;
- b) Instituições de pesquisa: Universidade Federal de Minas Gerais, Universidade Federal de Ouro Preto; Fundação João Pinheiro;
- c) Entidades ligadas ao setor produtivo: Banco de Desenvolvimento de Minas Gerais, Federação das Indústrias do Estado de Minas Gerais (FIEMG), Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE), Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAEMG) e empresa de saúde complementar. (PESSOA; FREITAS; BORGES, 2016, p.16-17)

De acordo ainda com os autores, o grupo técnico ampliado realizou, ao todo, nove reuniões, tendo oito delas ocorridas entre o primeiro semestre de 2012 e meados de 2013, nas fases de planejamento e concepção da plataforma, e uma reunião após a entrega da primeira versão, em outubro de 2013. (PESSOA; FREITAS; BORGES, 2016)

“Na fase de planejamento, o escopo das reuniões do Grupo esteve ligado à discussão dos elementos necessários ao delineamento do conteúdo da ferramenta em desenvolvimento, o que consistiu em: alinhamento e apresentação da teoria econômica do “*product space*”¹⁰ a todos os envolvidos; tratamento das bases de dados; apresentação de conceitos de gestão do conhecimento; apresentação de protótipo e resultados parciais do desenvolvimento; apresentação de metodologias de pesquisa econômica e prospecção ativa de investimentos; discussão das funcionalidades e usos da plataforma.” (PESSOA; FREITAS; BORGES, 2016 p.17)

¹⁰ Hidalgo et al (2007) desenvolveram um modelo de análise econômica conhecida como “Espaço de Produto” (*Product Space*), a partir da “Teoria das Capacidades” (SEN, 1993; 2000), na qual buscam responder à questões relacionadas ao crescimento e desenvolvimento da economia nos países. Segundo Pessoa, Freitas e Borges (2016), o *Product Space* é uma metodologia que mostra quais são os produtos que requerem habilidades, conhecimentos e tecnologias similares àquelas de que a localidade já dispõe, ajudando a desenhar uma política pública industrial e tecnológica orientada pelos próprios dados econômicos.

6 CONSIDERAÇÕES FINAIS

À época em que estava concluindo e finalizando este trabalho, foi publicada notícia, na página oficial da plataforma DataViva, com os dizeres que o “DataViva (estaria) de casa nova” (Vide Anexo I). Segundo as informações da página, a partir de 2017, o DataViva deixa a FAPEMIG e passa a integrar o Instituto de Desenvolvimento Integrado (INDI). Este fato, embora superveniente, tem bastante relevância e merece um pequeno destaque dentre as considerações que serão tecidas neste capítulo final, dados os seus impactos e reflexos acerca dos usos e potencialidades da plataforma DataViva.

O INDI é a agência pública local (do Governo de Minas Gerais) de promoção de investimentos e comércio exterior em Minas. O Instituto realiza consultoria a investidores, e gestores públicos, apoiando a prospecção e o desenvolvimento das empresas instaladas no Estado, e auxiliando as empresas mineiras que querem exportar ou importar. Ademais, o INDI presta uma extensa gama de serviços, a entidades privadas e governamentais, fornecendo apoio e assessoramento técnico, conforme pode ser visto abaixo:

Figura 7: Serviços prestados pelo INDI (consultoria especializada nas áreas de:)



Fonte: <http://www.indi.mg.gov.br/o-indi/servicos-prestados-pelo-indi/>

Conforme consta na notícia, em anexo, no dia 21 de dezembro de 2016, foi assinado contrato de licenciamento do DataViva entre os presidentes da FAPEMIG e do INDI.

Como visto até então, a plataforma trata-se de um robusto projeto governamental de visualização de dados públicos, com grande foco em desenvolvimento econômico e na caracterização das localidades brasileiras a partir de seus dados, disponibilizando informações socioeconômicas estruturadas e organizadas em mais de um bilhão de possibilidades de visualizações, sendo “a maior plataforma aberta de dados disponível no Brasil”.

É até difícil não pensar em termos de *Big Data*, com enfoque específico no apoio à tomada de decisão e formulação de políticas públicas, no caso do DataViva. Tendo isto em mente, a incorporação do DataViva pelo INDI surge neste momento, e segundo a própria matéria, como uma estratégia de ampliação de sua capacidade analítica, voltada à atração de investimentos (que hoje é o principal negócio do INDI), a partir do desenvolvimento de estudos técnicos com as diversas informações disponíveis na plataforma.

De acordo com o seu plano estratégico de sustentabilidade, desenvolvido no segundo semestre de 2016, o DataViva deverá receber novos direcionamentos, agora sob gestão do INDI. De acordo com o Coordenador do Projeto na FAPEMIG, Thiago Borges, “a plataforma sempre foi orientada para atingir os mais diversos públicos (até mesmo por ser totalmente livre e aberta), mas constatou-se a necessidade de focar em públicos específicos, por exemplo, a partir de novas propostas de atendimento direcionado a prefeitos e gestores municipais, fornecendo auxílio na sua tomada de decisão”.

Segundo o texto, a passagem da gestão do projeto também deve trazer consigo novas parcerias e ampliar o valor da ferramenta, indo além das diversas visualizações possíveis e passando a entregar conhecimentos aplicados sobre desenvolvimento econômico, sobretudo para as administrações municipais e regionais, já que a plataforma dispõe de dados de todo o Brasil. Este talvez fosse justamente o seu maior potencial, ainda subutilizado até então e que agora sinaliza aprimoramentos futuros.

Segundo a própria equipe técnica do DataViva, e como é possível observar tendo utilizado e estudado a plataforma, o arranjo institucional e o

desenho colaborativo de concepção e desenvolvimento da plataforma, implicam em uma expressiva dominância de entidades atuantes a nível local ou regionalmente, ou seja, com alcance e visibilidade restrita apenas ao estado de Minas Gerais.

Observando novamente, conforme discutido no decorrer do estudo, que o DataViva é uma plataforma que contempla dados para todos os municípios brasileiros, essa visibilidade restrita faz com que seu potencial de uso seja apenas parcialmente aproveitado, e implica em um desafio específico para o DataViva que é a sua divulgação, disseminação e uso pelas diversas regiões do país. Mais ainda, segundo a própria equipe, como pode ser visto na matéria, esse é um desafio que passa também pelo fortalecimento de uma cultura de dados e do uso de evidências empíricas na formulação de políticas públicas e na tomada de decisão organizacional - pública e privada.

Conforme visto também no decorrer deste estudo, o uso adequado das Tecnologias de Informação e Comunicação, de modo amplo, traz consigo grandes possibilidades de aprimoramento, sistematização e compreensão dos dados e informações que as organizações possuem e, poderia, em grande medida, permitir à Governos e à Gestão Pública, a automatização e aprimoramento de fluxos de informações internos e externos, dispondo de mais e melhores recursos para tomar decisões acertadas e desenvolver estratégias bem-alinhadas com a realidade.

Entende-se com isso que, no setor público, gerir a informação adequadamente torna-se ainda mais relevante \ importante, pois pode culminar na eficiência e efetividade dos serviços prestados à sociedade e na formulação de melhores políticas públicas.

Para ilustrar essa importância da gestão adequada da informação, no âmbito da tomada de decisão do setor público, imagine-se o seguinte exemplo: a informação, em si e por si só, é fundamental para o funcionamento das economias. Tomemos por exemplo, as informações relativas aos preços. Os preços são essenciais na coordenação das atividades econômicas, pois

forneem uma importante informaão sobre os desejos e necessidades por bens e servios.

Ponderando sobre o papel da informaão em contexto macroeconômico, sabe-se que determinado investimento inicial, pblico ou privado, no promove um crescimento de forma homognea, constante ou equilibrada. Na realidade, o crescimento tenderia a concentrar-se, em curto prazo, em determinados ramos da economia, diretamente relacionados ao investimento inicial e, so ento, difundir-se-ia para outros setores, mediante aumento de produão e/ou avanos tecnolgicos.

Assim sendo, a identificaão de setores estratgicos para receber investimentos torna-se de suma importncia no sentido de possibilitar a melhor escolha de investimentos, em setores capazes de provocar e atrair tambm novos investimentos, criando-se assim um ciclo de crescimento e desenvolvimento econmico. Da o enorme potencial de aplicaões para gestores pblicos de conhecerem e aplicarem informaões socioeconmicas de seu territrio na sua gesto, com o uso aplicado do DataViva.

Vimos tambm, que processos organizacionais de desenvolvimento de produtos (sofisticados ou complexos) de informaão (ainda mais tratando-se de iniciativas de *Big Data*) envolvem inmeros tipos de dificuldades e desafios, que vo desde o levantamento de todas as informaões a serem coletadas e supridas, definião de uma escala de prioridades para a gesto informacional, abertura de canais de divulgaão e veiculaão das informaões, motivaão e uso interativo das informaões (simples) pelos usurios, sistemas de *feedback*, formaão de equipes para viabilizar o(s) sistema(s) de informaões, racionalizaão da aplicaão dos recursos, entre outros.

Alm de tudo isso,  tambm um fator essencial no desenvolvimento dessas e de quaisquer aplicaões e produtos de informaão, a anlise das necessidades de informaão dos usurios.

Tendo em vista os seus pblicos-alvo, os critrios de transparncia e livre acesso, o seu enfoque econmico e a possibilidade de obter diversas

visualizações distintas para a análise e estudo dos dados, pode-se dizer e concluir que o DataViva constitui um robusto “produto informacional” desenvolvido através de um arranjo colaborativo e governamental específico, e de investimentos na qualidade e na gestão da informação de bases de dados públicas.

Ao apontar novas relações e evidenciar dinâmicas econômicas que não eram vistas ou associadas até então, a plataforma permitiria a atores públicos e privados o aprimoramento de suas estratégias e decisões. De tal sorte, mais do que disponibilizar informações públicas, o DataViva é uma plataforma colaborativa, em que todos podem explorar seus dados, de maneira livre, crítica ou mesmo propositiva. A plataforma, que invariavelmente constitui uma iniciativa governamental de *Big Data*, (vide Anexo II), a despeito de possíveis diferenças conceituais, possibilita conexões e relações antes impossíveis entre os diferentes dados, possibilitando compreender aspectos do desenvolvimento social e do desenvolvimento das diversas atividades econômicas formais no país.

Proporcionando também o acesso livre e gratuito a todas as suas informações e até mesmo à própria estrutura que compõe a plataforma, para que outros governos possam adaptá-la de acordo com suas necessidades, o DataViva disponibiliza um conhecimento técnico, específico e detalhado de diversos fatores socioeconômicos brasileiros, a partir de bases de dados públicas, provendo insumos para o planejamento e a tomada de decisão por parte de empresários, estudantes, investidores e profissionais de qualquer setor econômico. Trata-se de uma iniciativa bastante criativa que renova a governança pública e colaborativa, fortalece uma cultura de dados públicos abertos e transparência e é altamente aplicável à tomada de decisões.

Deste ponto de vista, nossa opinião acerca da plataforma apresentada neste estudo de caso é extremamente positiva. Espera-se que esta plataforma possa continuar evoluindo e trazendo bons frutos àqueles que a utilizam.

Toda essa discussão em torno do DataViva, no entanto, acaba tirando um pouco o foco de outro aspecto também fundamental a ser ressaltado no

âmbito destas considerações finais, tratando-se ainda acerca do *Big Data*: *Big Data* é (antes de tudo) sobre os dados.

Antes de se definir metas e objetivos para iniciativas de *Big Data*, é fundamental que seja realizado um inventário das informações disponíveis, porque são elas que irão determinar o que é possível ou não de ser feito. Algumas vezes, uma simples combinação de dados externos com dados e informações que a organização já possui pode significar um enorme aumento no seu valor, mesmo sem a aplicação de qualquer tipo de análise sofisticada ou complexa. Por outro lado, e conforme visto ao se tratar de aspectos relacionados à Qualidade de Dados, nem as melhores e mais modernas ferramentas podem extrair valor de um conjunto ruim de dados.

Há muito tempo (algumas décadas, talvez) já existem diversas organizações que trabalham com o que hoje em dia se convencionou em chamar de *Big Data*. Empresas de telefonia, por exemplo, bancos e financeiras, e mesmo empresas de internet, como Google e Amazon, sempre trabalharam com robustos conjuntos de informação, com grandes volumes, variedade e velocidade em processamento e análise.

Ainda, há pouco mais de dez ou quinze anos atrás não existia o aporte tecnológico e científico de que dispomos hoje, e o que torna o caso dessas organizações citadas acima relevante, e que, de fato, motivou o desenvolvimento de novas tecnologias e ferramentas, foi justamente o **valor** que estaria contido nos dados de que elas dispunham.

Assim, antes de se discutir qualquer iniciativa ou projeto de *Big Data* deve-se sempre entender a fundo a natureza dos dados e informações disponíveis e planejar adequadamente tudo que pode ser extraído deles. Algumas vezes, o maior valor de um conjunto de informações não está nem mesmo relacionado à sua aplicação original. É importante frisar que muitos projetos de *Big Data* surgem a partir de se imaginar novas aplicações para dados que já existem, especialmente quando relacionados com outras informações que estão espalhadas por aí e podem ser coletadas e analisadas conjuntamente.

Posto isso, constata-se também que para a governança colaborativa e a gestão compartilhada das informações entre diferentes organizações, é fundamental a presença de um contexto organizacional propício e orientado ao desenvolvimento de estratégias para o desenvolvimento, assim como a inclusão da inovação na agenda governamental e o envolvimento ativo e transparente do Estado na gestão e no compartilhamento da informação, visando, em última análise, ao aprimoramento das estruturas e capacidades do Estado.

Assim, o valor contido, e, portanto, passível de extração, dos dados e os arranjos necessários para a sua gestão adequada, convergem para um contexto em que as tecnologias e ferramentas utilizadas para se trabalhar com o *Big Data*, constituem um fator secundário ou de menor importância, de modo que nem foram amplamente discutidas no decorrer deste estudo, pois estas tenderão a ser cada vez mais simples, interativas e eficazes, em termos de custo-benefício, perante as organizações, e seus usuários.

Neste contexto, é simples perceber o que aqui se conclui: *Big Data* é (antes de tudo) sobre os dados.

Referências

ASSIS, Wilson Martins de. **Metodologia para construção de produtos de informação nas organizações**. 2006. 251f. Dissertação apresentada ao curso de Mestrado em Ciência da Informação – Escola de Ciência da Informação – Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

BARBIERI, Carlos. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. Fumsoft - Belo Horizonte, 2013.

BRESSER-PEREIRA, Luiz Carlos, **Administração Pública Gerencial: estratégia e estrutura para um novo Estado**. Brasília: MARE/ENAP, 1998.

CAMPOS, Fábio Rocha. **A gestão da inovação em serviços intensivos em conhecimento: oportunidades e desafios do Big Data**. 2015. 124f. Dissertação apresentada ao curso de Mestrado em Geociências - Universidade Estadual de Campinas, Campinas, 2015.

CANARY, Vivian Passos. **A tomada de decisão no contexto do Big Data: Estudo de caso único**. 2013. 74 f. Trabalho de Conclusão de Curso de Graduação (Administração)- Escola de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/87757/000911900.pdf?sequence=1>>. Acesso em: 10 out. 2016.

COSTA, Sônia Regina Neves. Dados Abertos e sua utilização por pesquisadores de Comunicação Social **Brazilian Journal of Technology, Communication, and Cognitive Science** c Edição nº 1, Ano I - Setembro 2013

GOMES, Eduardo Granha Magalhães. **Gestão por Resultados e eficiência na Administração Pública: uma análise à luz da experiência de Minas Gerais**. São Paulo: EAESP/FGV, 2009, 187 p.

GUERRA, Alexandre Lalau et al. **Diagnóstico e Formulação de Políticas Públicas em Municípios com menos de 50 mil habitantes: Recomendações e Instrumentos**. 2012. 130 p. Dissertação (Mestre em Gestão e Políticas Públicas)- Fundação Getúlio Vargas, São Paulo, 2012. 1. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/10047/%28MPGP_P_CEPAM_Vers%c3%a3o%20Final_22_09_12%29.pdf?sequence=1&isAllowed=y>. Acesso em: 09 nov. 2016.

LAKATOS, Eva Maria; MARCONI, Maria de Andrade. **Fundamentos da metodologia científica**. São Paulo: Editora Atlas, 2010.

MAFLI, Luana Michele de Souza. **Gestão da Informação e Gestão do Conhecimento: Uma análise desses processos nas organizações**,

evidenciando sua implementação no contexto da SEPLAG-MG. 2015. 65 f. Monografia apresentada ao Curso (Especialização em Gestão de Informação e Pessoas)- Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

DE PAULA, Claudio Paixão Anastacio. **Novas tecnologias, gente, informação, perda de sentido e o retorno da alma**: alguns tópicos para a compreensão das percepções de tempo e espaço na pós-modernidade. Disponível em: <http://media.wix.com/ugd/731c19_4d43ba79c046042caafb6bca6ad0c415.pdf> Acesso em jun.2016.

PESSOA, Rafael Marques; FREITAS, Elton Eduardo; BORGES, Thiago Bernardo. **DataViva**: Plataforma de Visualização de Dados Públicos Socioeconômicos Brasileiros. In: Congresso CONSAD de Gestão Pública, IX., 2016, Centro de Convenções Ulysses Guimarães, Brasília. Painel 35/001 **Dados Abertos e Novas Perspectivas de Acesso à Informação pelo Cidadão**. [S.l.: s.n.], 2016. p. 2-22. Disponível em: <<http://www.dataviva.info/pt/scholar/article/67>>. Acesso em: 07 nov. 2016.

RIBEIRO, Claudio Jose Silva; ALMEIDA, Reinaldo Figueiredo de. **Dados Abertos Governamentais (Open Government Data)**: Instrumento para exercício de cidadania pela sociedade. 2011. XII ENANCIB. Encontro Nacional de Pesquisa em Ciência da Informação. Brasília. 2011 disponível em: http://eprints.rclis.org/18019/1/GT_8.pdf#page=124

TAURION, Cezar. **Big Data**. Rio de Janeiro. Brasport, 2012.

SEBRAE, **Formulação de políticas públicas**, 2014. Disponível em: <http://www.sebrae.com.br/sites/PortalSebrae/artigos/formulacao-de-politicas-publicas,e38b9e665b182410VgnVCM100000b272010aRCRD>

SIDNEY, Ana Luiza Gusmão. **Acompanhamento da Proposta e das Ações de Governança de Qualidade de Dados do Projeto MASP**. 2015. 110 f. Monografia de Conclusão de Curso de Graduação (Bacharelado em Administração Pública)- Escola de Governo Professor Paulo Neves de Carvalho , Fundação João Pinheiro, Belo Horizonte, 2015.

WORLD ECONOMIC FORUM. **Big Data, Big Impact**: New Possibilities for international development. Cologny, Geneva- Switzerland. 2012.

ANEXO I:

DataViva de casa nova

por Equipe DataViva 22/12/2016

<http://www.dataviva.info/pt/news/publication/33>

Notícias

DataViva de casa nova

POR EQUIPE DATAVIVA 22/12/2016



A partir de 2017, o **DataViva**, projeto de visualização de dados públicos com foco em desenvolvimento econômico, passa a integrar o **Instituto de Desenvolvimento Integrado (INDI)**. Nesta quarta-feira, 21, foi assinado contrato de licenciamento entre os presidentes do INDI e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Fapemig, onde está alocado o projeto. A plataforma, que disponibiliza dados socioeconômicos nacionais estruturados em mais de um bilhão de possibilidades de visualizações, foi desenvolvida em parceria entre o Governo de Minas Gerais e o MIT Media Lab, com financiamento da Fapemig.



Da esquerda para a direita, Prof. Evaldo Vilela, presidente da Fapemig, Cristiane Serpa, diretora-presidente do INDI, Olavo Machado Junior, presidente da FIEMG e José Roberto Scolforo, reitor da Universidade Federal de Lavras, na transferência do DataViva para o INDI.

“Nesses últimos dois anos, a Fapemig deu uma cara brasileira para o DataViva, e queremos que a ferramenta contribua para uma mudança de paradigma, para que decisões sejam tomadas com base em dados”, afirmou o presidente da Fapemig, Prof. Evaldo Vilela. Sob gestão da Fapemig, o DataViva passou por um período de recuperação, com reestruturação da equipe e da própria plataforma, que ganhou sua versão 3.0, mais focada no público nacional. “Agora o projeto está pronto para evoluir, com uma equipe constituída e um plano estratégico de execução, pronto para dar novos passos, melhorar a tecnologia e o acesso. O desafio é que, quando se pensar em dados de desenvolvimento econômico, a plataforma apareça como referência, e passe a produzir conteúdo relevante para a tomada de decisão”, diz Thiago Borges, coordenador do projeto na Fapemig.

Agência de promoção de investimento do estado, o INDI incorpora o DataViva como estratégia de ampliação da capacidade analítica voltada à atração de investimentos, a partir do desenvolvimento de estudos com informações disponíveis na plataforma. “Nós acompanhamos o DataViva desde que foi criado, e agora vamos ter a chance de desenvolver a ferramenta internamente, a partir de orientações da Fapemig”, conta Letícia Vargas, coordenadora do projeto no INDI. A integração está prevista para o início de 2017, com a migração da equipe do projeto para o Instituto, e após um trabalho de reposicionamento estratégico realizado em parceria com a consultoria Aleph.

O plano estratégico de sustentabilidade, desenvolvido no segundo semestre de 2016, deve conferir rumo aos novos passos do projeto, agora sob gestão do INDI. “O DataViva buscou atingir vários públicos, mas percebemos que precisávamos focar em um público específico, e visualizar uma proposta de atendimento aos gestores municipais, ajudá-los na tomada de decisão”, explica Thiago Borges. A passagem do projeto deve gerar novas parcerias e ampliar o valor da ferramenta, indo além das visualizações e passando a entregar conhecimentos sobre desenvolvimento econômico, sobretudo para as administrações municipais.

“O DataViva é uma plataforma muito completa, com espaço para incluir mais informações. E a importância dela está em transformar esses dados em conhecimento. Mas temos a impressão de que é subutilizada, e acreditamos que dentro do INDI a ferramenta terá uma aplicação mais direta”, afirma Letícia Vargas. A incorporação do projeto ocorre junto à integração do Exportaminas ao INDI. A expectativa é que, junto ao Exportaminas, o DataViva ajude a identificar a produção, a exportação, a concorrência e vazios na cadeia produtiva do estado, auxiliando na alocação e atração de investimentos públicos e privados.

ANEXO 2: TEXTO TINNO COMUNICAÇÃO

AFINAL, O QUE É *BIG DATA*



Afinal, o que é big data?

Para que, como funciona e quais os usos do big data

Texto Tinno Comunicação

Afinal, o que é big data?

Introdução



Muita gente tem uma ideia remota do que seja o big data, afinal, já faz tempo que o termo corre solto em conversas no meio empresarial. Ainda assim, é comum que as pessoas tenham dúvida sobre o que é, realmente, essa tão revolucionária solução.

A verdade é que o big data não é nem uma ferramenta, nem um conjunto de informações, nem uma técnica ou serviço: esse termo faz referência a um fenômeno em que o volume de informações disponíveis cresceu muito e muito rápido, tornando necessária a criação de novas plataformas de análise para processar essa quantidade imensa de dados.

2

Índice



A era do petabyte	4
Ok, mas para que serve tudo isso?	5
Quer ver como funciona?	6
E você também pode usar	9

A era do petabyte



3,2 bilhões: esse é o número estimado de usuários de internet no mundo inteiro. Mais impressionante ainda é pensar que, há 15 anos, esse número era de apenas 400 milhões. Considerando que tudo o que fazemos no mundo virtual gera informação – cada clique, cada palavra pesquisada, cada conteúdo criado –, fica fácil entender que o volume de dados disponíveis andou crescendo drasticamente.

Para se ter uma ideia, as empresas que trabalham com análise desse mundo de informações utilizam a medida petabytes – que equivale a mil terabytes ou, ainda, um milhão de gigabytes. Difícil de imaginar? Aqui vão algumas estatísticas comparativas para tornar as coisas mais claras:

- 1 petabyte equivale a 13,3 anos de vídeo no formato HD-TV;
- 10 bilhões de fotos do Facebook ocupam 1,5 petabyte;
- O Google processa 20 petabytes por dia;
- 50 petabytes é tudo que a humanidade escreveu na história, desde que se tem registro, em todas as línguas.

O tratamento estatístico tradicional já não dá conta de analisar tanta coisa. Por isso, novas plataformas vêm sendo criadas para capturar, estruturar e analisar toda essa informação. E é a todo esse fenômeno de geração e de análise de dados que se dá o nome de big data.

Afinal, o que é big data?

Ok, mas para que serve tudo isso?



Já não dá mais para negar que a internet é uma fonte infinita de conhecimento útil. De horário de sessões de cinema a números que indicam qual o melhor investimento a ser feito por uma empresa, a web provê informações que impactam na vida das pessoas e, em um nível mais amplo, no desempenho da economia e na geração de produtos e conhecimentos.

A análise de big data pode, então, reunir e organizar todo esse conteúdo caótico e transformá-lo em dados inteligíveis e aplicáveis. Assim, pesquisadores podem ter acesso a informações sobre padrões de comportamento a dados que auxiliam no desenvolvimento de novas tecnologias; empresas podem avaliar suas atividades ou criar estratégias de vendas; e até um público menos especializado, como estudantes e profissionais de qualquer área, pode se aventurar em ferramentas que oferecem dados como empregos disponíveis e renda por ocupação, por exemplo.

5



Afinal, o que é big data?

Quer ver como funciona?



O mais legal dessa tecnologia é que ela pode ser utilizada para tantas finalidades que fica difícil enumerar. Aqui vão alguns exemplos de uso do big data para situações como aumento de vendas, desenvolvimento de produtos e personalização de serviços.

Big Data para aumento de popularidade: o caso dos sites Huffington Post e FT.com

Em um cenário online que cresce e muda rapidamente, o Huffington Post conseguiu se tornar o site mais acessado nos Estados Unidos. De acordo com pesquisa feita pelo Reuters Institute, a liderança da companhia acredita em gerir o negócio com base em big data. Isso inclui melhorar a experiência do usuário em tempo real a partir de tendências sociais, recomendações e personalização, otimizando o site de várias maneiras.

6

O site do jornal Financial Times também utiliza big data para entender como oferecer um serviço melhor ao seu consumidor, criando publicidade segmentada e fazendo o design de novos produtos, baseado em informações coletadas. O CEO da empresa afirmou que o big data transformou o negócio deles. A companhia utiliza vários pontos de dados para analisar as preferências do leitor e oferecer um conteúdo personalizado – tudo isso para aumentar o tráfego do site. Os dados também ajudam a organização a entender o período do dia em que seu conteúdo é acessado e se as visitas estão vindo de desktops ou de aparelhos móveis.

Afinal, o que é big data?



Big data para desenvolvimento de produtos: o caso do McDonald's

O McDonald's coleta e combina os múltiplos dados de suas lanchonetes ao redor do mundo a fim de padronizá-los e, com isso, compreender as reações de seu público, as expectativas de cada nicho em torno de seus produtos e as alterações logísticas e de design que podem ser feitas para melhorar a cadeia de serviços.

Já foram lançados novos sanduíches a partir de estudos de Sentiment Analysis (análises de sentimento) realizados em mídias sociais; foram feitas promoções em tempo real, acompanhadas de perto por cientistas de dados, que mensuraram atentamente as manifestações e reações de seu público, alterando estratégias em tempo real; e até a logística do drive-thru é alterada em cada país de acordo com as reações de seus consumidores no que concerne a questões como design, tempo de espera e informações providenciadas por seus funcionários no ponto de retirada dos produtos. Tudo por meio de ferramentas de Big Data Analytics.

7

Afinal, o que é big data?



Big data para aperfeiçoar o serviço: o caso da Tesla e da Ford

Tesla foi a primeira companhia que acoplou em seus veículos sensores que enviam dados a um sistema central. Esses dados são utilizados para melhorar a performance do veículo e analisar a satisfação do cliente. A empresa é notificada se um carro não está funcionando corretamente e os clientes podem receber instruções de como resolver o problema. Essas capacidades têm ajudado a Tesla a criar uma expressiva cota de mercado e torná-la um caso de sucesso no uso de big data.

A Ford tem seguido o mesmo caminho e buscado utilizar big data para melhorar o serviço oferecido a seu cliente. A empresa quer usar dados para diminuir o custo de seguro e fazer carros que respondam às expectativas dos clientes. Para isso, ela tem aproximadamente uma equipe de 200 experts, que analisa big data e influencia em grandes decisões da empresa. Além disso, a Ford analisa o que os consumidores estão procurando em websites e o perfil socioeconômico de potenciais compradores – tudo isso para oferecer uma demanda que condiz com a procura.

8

Afinal, o que é big data?

E você também pode usar!



O cerne da análise de dados está em encontrar insights capazes de influenciar decisões. Mas isso não quer dizer que apenas grandes empreendedores e CEOs podem fazer uso de big data, afinal, dados podem ter as respostas para dúvidas que surgem no dia-a-dia de estudantes, pesquisadores e agentes do governo.



9

Afinal, o que é big data?

O DataViva, desenvolvido pela Fapemig, é uma plataforma de pesquisa aberta que conecta os usuários a mais de 1 bilhão de visualizações de dados socioeconômicos sobre, aproximadamente, 5 mil municípios brasileiros. A ferramenta é um exemplo de big data sendo utilizado para soluções práticas do cotidiano. Uma vez que as informações fornecidas perpassam uma gama variada (atividades econômicas, educação superior, cursos básicos, produtos, parceiros comerciais), é possível realizar pesquisas bem direcionadas para perguntas específicas.

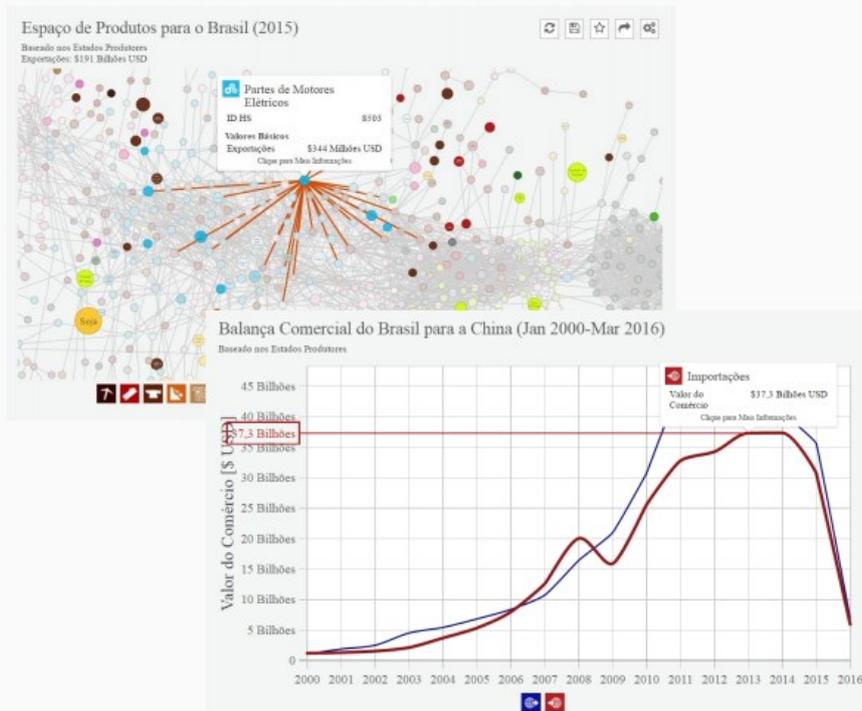


Se você achava que precisaria de softwares complicados para o utilizar o big data a seu favor, está enganado. O DataViva torna o acesso a uma quantidade gigantesca de dados algo simples e intuitivo. Além disso, a ferramenta possibilita que cada indivíduo procure uma correlação entre os dados – o que é crucial, pois entender indicadores a partir de sobreposições de camadas pode ser ainda mais esclarecedor.

10

Localidades Brasileiras Perfil geral por região, estado, mesorregião, microrregião ou município brasileiro. Visualize dados de comércio exterior, atividade econômica, empregos e educação.	Ocupações Regiões que mais empregam por atividade profissional, cursos relacionados, salário médio e estatísticas de emprego por ano.	Atividades Econômicas Informações sobre taxa de empregos por região, salário médio por tipo de ocupação, renda mensal média e oportunidades econômicas.
Produtos Dados da Balança Comercial por produto, origem das importações e destino das exportações, ranking por localidade, atividades econômicas e ocupações relacionadas.	Parceiros Comerciais Nações que importam ou exportam produtos para o Brasil, dados de comercialização por tipo de produto ou por município comercializador.	Universidades Número de matrículas em cada curso ofertado, situação dos estudantes e perfis semelhantes de universidades.
Ensino Superior Lista de universidades e municípios que ofertam o curso de Educação Superior ou o campo de estudo selecionado.	Cursos Básicos Escolas do Brasil que oferecem educação profissional e ranking de matrículas na Educação Básica por município.	

Afinal, o que é big data?



11

Afinal, o que é big data?



Se você é um empreendedor, por exemplo, pode fazer uma pesquisa com foco em oportunidades de negócio referentes a áreas específicas. Se você está à procura de emprego, o big data pode ser seu melhor amigo para descobrir dados sobre oportunidades e valores de salário ao redor do Brasil. E, ainda, se você é um agente público, compreender as características econômicas e de mercado de trabalho de municípios ou regiões pode ajudar na tomada de decisões sobre ações públicas relacionadas.

Como podemos ver, em vários momentos em que precisamos tomar decisões, o embasamento focado em estatísticas assertivas pode ser um passo para o sucesso.

12

É isso!

Esperamos que, agora, para você, o big data tenha deixado de ser uma grande dúvida e se mostrado como realmente é: uma fonte de resposta para vários tipos de perguntas!

(FIM)