

INTEGRAÇÃO ABERTA PARA QUALIDADE DO CONHECIMENTO

DISTRIBUÍDO: Abordagem na Coleta de Dados Geoespaciais em Cidades Sustentáveis

André Fabiano de Moraes¹

Lia Caetano Bastos²

Fernanda dos Santos Koehler³

Marina Souza Kracik⁴



Neide Inês Grasel de Moraes⁵

Abstract: *The growing volume of spatial data, together with the diversity of non-integrated solutions and heterogeneous computational structures, make it difficult to generate knowledge and support decision. In parallel, there is a shortage of technologies that helping in the quality of information and structuring of strategies. This context, paper presents an approach to identify infrastructure and applicable frameworks in the implementation of Big Data. Providing engineering steps for knowledge ownership and better management of cities. For automatic data acquisition, web services are elaborated along with (IoT) Internet of Things devices, for greater adherence and quality of the information in processes. In conclusion, are discuss results achieved with the case study aiming at the generation and propagation of knowledge.*

Keywords: *integration, iot, quality, big data, knowledge.*

Resumo: *O crescente volume de dados espaciais aliado a diversidade de soluções não integradas e estruturas computacionais heterogêneas, dificultam a geração de conhecimento e o apoio à decisão. Paralelamente se tem carência por tecnologias que auxiliem na qualidade das informações e estruturação de estratégias. A partir deste contexto, o trabalho apresenta uma abordagem para identificar infraestrutura e frameworks aplicáveis na implantação de Big Data. Proporcionando a engenharia das etapas para apropriação do conhecimento e melhor gestão das cidades. Para coleta automática de dados, são elaborados web services juntamente com dispositivos de IoT (Internet das Coisas), para maior aderência e qualidade das informações em processos. Concluindo, são discutidos os resultados alcançados com o estudo de caso visando a geração e propagação de conhecimentos.*



Palavras-chave: *integração, iot, qualidade, big data, conhecimento.*

1 INTRODUÇÃO

A engenharia em sintonia com o futuro, sustentabilidade, tecnologia e inovação, para Melhado (2018) e Palandi (2018), este é o foco para instigar reflexões que irão contribuir para a formação dos profissionais envolvidos com projetos, a fim de quebrar o paradigma de que a sustentabilidade é onerosa e, portanto, voltada apenas para cidades de alto padrão. Também, a partir de desafios globais na busca de soluções, surgem novas oportunidades, no que diz respeito a constante geração de dados e a visualização de volumes expressivos de

1 Engenharia e Gestão do Conhecimento - UFSC - Florianópolis - Brasil. afdmoares@gmail.com

2 Engenharia e Gestão do Conhecimento - UFSC - Florianópolis - Brasil. liacbastos@gmail.com

3 Engenharia e Gestão do Conhecimento - UFSC - Florianópolis - Brasil. admfernandadosantos@gmail.com

4 Engenharia e Gestão do Conhecimento - UFSC - Florianópolis - Brasil. mahkracik@gmail.com

5 Engenharia Civil - Faculdade CNEC - Itajaí - Brasil. neideines@gmail.com

informações, além de ferramentas dinâmicas que auxiliem na otimização e disponibilização de diferentes fontes de dados geoespaciais através da rede internet (Cunha, 2018).

A crescente disponibilidade de dados e informações cria a necessidade de novos processos supervisionados e não supervisionados, a exemplo dos trabalhos de Simon et al. (2018), onde a extração de dados ganha espaço com novas demandas na geração de conhecimento, a partir de algoritmos, frameworks e softwares de ETL (*Extract Transform Load*), denominadas de *Big Data Analytics*.

Um dos diferenciais do *Big Data* se encontra na possibilidade de focar no processamento de dados em busca de correlações e descobertas. Nos últimos anos, as empresas têm se voltado principalmente para a existência de *Big Data* em relação a vários tipos de informações relacionadas a seu cotidiano (ambiente interno, concorrência, tendências, sociedade, dentre outras fontes de dados), um dos tipos de sistemas em evidência no mercado é o sistema de apoio à decisão (SAD).

Além do uso organizacional, as soluções de *Big Data* são utilizadas com foco nas cidades, avançando dos estudos chamados de “cidades digitais”, nos anos 90 (Lemos, 2013). Esses estudos, também contribuíram para o acesso da população às ferramentas de tecnologia da informação, utilizando a internet e a grande disponibilidade de dados de geolocalização, preferências de consumo, dentre outras, permitiram a criação de conhecimento que além de movimentar grande quantidade financeira, pode apresentar soluções a problemas comuns da sociedade.

O termo “cidades digitais” avançou então para “cidades inteligentes” (em inglês *smart cities*), sendo então “uma cidade na qual tudo é sensível ao ambiente e produz, consome e distribui um grande número de informações em tempo real” (Lemos, 2013, p.48). Nesse contexto, as pessoas atuam não apenas como receptoras de informação, mas também como produtoras da informação. Ainda, Kon e Santana (2016) & Montori et al. (2018), destacam como algumas áreas de aplicação e serviços para cidades inteligentes, o monitoramento do trânsito, condições da cidade, controle de multidões, sistema de coleta de lixo, sistemas de saúde, segurança pública, gerenciamento de recursos, riscos dentre outros.

Em relação ao gerenciamento de riscos, de acordo com Bertone e Marinho (2013), desastres naturais têm ocorrido com maior frequência no Brasil, sendo as inundações e enxurradas em maior número. Gerando para Hai et al. (2008) e Shen (2011) formas para uma análise quantitativa do padrão espacial. O estado de Santa Catarina possui histórico de desastres naturais com destaque para inundações e deslizamentos principalmente nas regiões que abrangem a bacia do Vale do Itajaí, que é a maior do estado de Santa Catarina

sendo que o Rio Itajaí-Açu é o de maior curso d'água. Conforme Silva (2016), a Bacia Hidrográfica do Itajaí, correspondente a 16,14% do território catarinense, onde estão assentados 53 cidades de três regiões: Vale do Itajaí, Norte e Grande Florianópolis.

O uso de soluções de *Big Data* para a tomada de decisão em situações que demandam gerenciamento de risco tem pesquisas em andamento e neste contexto, este estudo descreve estudos de casos de ocupação em áreas de risco, as quais, conforme Buffon & Paz (2017), estão propensas aos desastres naturais. Destaca-se a bacia hidrográfica do Vale do Rio Itajaí no estado de Santa Catarina - Brasil, sendo escolhida como estudo de caso no qual foram realizados levantamentos para estruturação de *Big Data* que visa aperfeiçoar o monitoramento das características espaciais (Meirelles, 2018) e a melhoria da gestão do conhecimento.

A seguir o artigo pretende esclarecer e ampliar alguns aspectos conceituais para aquisição e integração de dados através de padrões e contribuir no ciclo de gerenciamento do conhecimento distribuído conforme Schreiber *et al.* (2000, p.71). Sendo pouco abordado na literatura pelo conceito de integração aberta como papel das organizações na construção e na economia do conhecimento. A partir deste cenário gerando oportunidades, que são favoráveis à manifestação de ideias e ao surgimento de conceitos e teorias, bem como à disseminação de novas tecnologias, entre elas, a Internet das Coisas (IoT) e *Big Data*. É proposto o levantamento teórico visando nortear ações para estender um ambiente interoperável e fomentar a qualidade das informações. Na sequência são propostos procedimentos metodológicos desde a coleta de dados, modelagem do sistema, estudo de caso, análise e discussão dos resultados alcançados.

2 INTERNET DAS COISAS (IoT)

A ideia de dispositivos conectados lembra a década 70, naquela época, se costumava ser chamada de “internet embutida” ou “computação generalizada”. Conforme Zanella (2014) ganhou outro termo “Internet das Coisas” e foi cunhado por Kevin Ashton, em 1999. Contudo, o consumo de energia para funcionamento desses dispositivos para essa tecnologia é um fator predominante a décadas. Ainda, o termo *IoT (Internet of Things)*, remete a dispositivos capazes de comportar tecnologias, que tenham conectividade com a internet, de forma que elas se comuniquem com outras tecnologias mutuamente. Esses dispositivos, referem-se a todo e qualquer aparelho que possa se conectar, desde os mais óbvios como os smartphones, videogames, notebooks e Smart TVs, até câmeras de segurança, veículos, sensores diversos dentre outros. A internet sem dúvida transformou a indústria e

os profissionais, trazendo tecnologias capazes de contribuir com seu ambiente organizacional. Sobretudo, agregando valor a serviços e negócios, com a melhoria de estratégias, tempo de resposta e redução de custos.

A *IoT* industrial e a indústria 4.0 juntas caracterizam um marco evolutivo. Por esse motivo, a automação e os robôs já alcançaram espaços significativos. A *IoT* industrial é simplesmente a conexão entre dispositivos no chão de fábrica, permitindo o desenvolvimento de sistemas cyber-físicos e comunicação entre dispositivos. Isso permite novos caminhos para gerar e coletar dados interessantes ao longo do espaço industrial. Também foca em dispositivos de seleção que proporcionam visibilidade até os pontos mais baixos de produção, fornecendo visibilidade na máquina e possibilitando a manutenção preventiva e análises de *Big Data*. A *IoT* industrial é, intrinsecamente, sobre conexão de dispositivos. Por outro lado, a indústria 4.0 é a fusão de mídias digitais, novas tecnologias, e decisões práticas focadas na mudança sobre os processos de fabricação de produtos através de flexibilidade, eficiência e visibilidade em cada nível da produção. Tem impacto ao longo dos processos e através da cadeia de mantimentos. Ao simplificar os processos, automação, tecnologia, materiais, redução do tempo de inatividade e bom uso de equipamentos em geral, a filosofia da indústria 4.0 revitaliza a competitiva no mercado global.

3 BIG DATA

É importante analisar as causas e os fatores de implantação do *Big Data*, auxiliando na compreensão de políticas para o processo de decisão (Taurion, 2013). E para melhor direcionamento das discussões abordadas neste trabalho, a partir dessas abordagem, é permitido avaliar as condições mínimas para o suporte à tomada de decisão nos moldes “tradicionais” atende às expectativas e às necessidades de um tomador de decisão no contexto do *Big Data*, bem como avaliar o grau de efeito dos fatores envolvidos na decisão abordando a grande quantidade de dados, a velocidade, a diversidade, valores ponderados e a confiabilidade.

O conceito de BI (*Business Intelligence*) é traduzido como “Inteligência de Negócio” e consiste na possibilidade das pessoas acompanharem tudo o que ocorre no negócio, assim como também no mercado. Pensando visualmente nos dados, são acessados por meio de tabelas dinâmicas, análise de performance ou painéis interativos, a exemplo dos dashboards. São fontes de dados que mostram o desempenho da equipe em produtividade, com panorama das gestões internas de pessoas e materiais. O diferencial do BI dispõe é o de facilitar o acesso a dados e informações exatas para guiar a tomada de decisão.

Por outro lado o *Big Data*, conforme Ramos (2015) e Marquesone (2017), teve expressivo avanço da tecnologia e tem permitido que incontáveis dados sejam trocados entre as pessoas e sistemas por meio de diferentes ferramentas. Quando estas trocas apresentam um grande volume de dados variáveis, que podem ser analisado e processados em alta velocidade, toda essa carga de informação recebe o nome de *Big Data*. Os dados podem ter origens distintas, e são classificados como estruturados, por exemplo dados catalogados e separados em planilhas e não estruturados (como se fossem “novos” de alguma análise; dados brutos). Ao processo de refino e análise por algoritmos de alta capacidade processual, se tem a denominação de *Big Data Analytics*. É justamente neste aspecto que se encontra o diferencial em *Big Data*: a possibilidade de focar no processamento de dados em busca de correlações e descobertas. Nos últimos anos, as empresas têm se voltado principalmente para a existência de *Big Data* em relação a vários tipos de informações relacionadas a seu cotidiano ambiente interno, concorrência, tendências, sociedade e dentre outras fontes de dados.(Malaquias *et al.*, 2017).

Ainda conforme Berman (2013) e Schreiner (2016), *Big Data* é uma denominação para um conjunto de dados que podem ser definidos com os chamados 3Vs: (i) Volume: são repositórios com grandes quantidades de dados, (ii) Variedade: diz respeito a dados com heterogeneidade de representação, e (iii) Velocidade: diz respeito a dados que são gerados com grande frequência e podem estar em constantemente evolução. O *Big Data* introduz desafios em termos de gerenciamento de dados, uma vez que os convencionais sistemas gerenciadores de bancos de dados são ineficazes para o seu tratamento (Kumar *et al.*, 2014). Ainda Klein & Gorton (2015) e Sun & Du (2017), para suportar os 3Vs, os sistemas de banco de dados que lidam com *Big data* devem ser distribuídos pautados no paradigma da nuvem, escaláveis, visando tratar requisições de leitura e escrita em paralelo.

3.1 QUALIDADE DOS DADOS E DA INFORMAÇÃO

A qualidade de dados é uma das áreas de conhecimento abordadas pela governança de dados na estrutura fornecida pelo Guia DMBOK® - *Data Management Body of Knowledge* (Dama, 2012; Barbieri, 2013). Ainda conforme Britto e Almeida Júnior (2006), define-se qualidade de dados como o grau de aderência entre as visões apresentadas pelos dados armazenados e os mesmos dados no mundo real. Ou seja, considerando o dado como uma representação de algum elemento do mundo real, quanto melhor for essa representação (mais próxima da realidade), melhor será a qualidade do dado. De acordo com (Klein, 2015), a qualidade de dados é definida na literatura como um conceito multidimensional,

cujas dimensões mais citadas são: precisão, completude, consistência, e frequente atualização, conforme no quadro 1.

Porém, Fagundes *et al.* (2017, p.4), descreve que um dado/informação tem qualidade se satisfaz os seguintes requisitos de uso: credibilidade, livre de erros, objetividade, reputação, representação concisa, representação consistente, interpretabilidade, compreensividade, acessibilidade, facilidade de manipulação, segurança, quantidade apropriada de dados, completude, relevância, tempestividade, e valor agregado. Ainda, a percepção dessas dimensões pode variar em função do sujeito que as analisa e serem mais ou menos importantes pelo uso da informação para um determinado tempo e espaço. Ao avaliar a qualidade da informação, o usuário poderá considerar todas as dimensões ou aquelas que melhor for conveniente, conforme Quadro 1. Para Assis e Moura (2011), na literatura sobre qualidade da informação encontram-se duas abordagens distintas: positivista, que utiliza bases empíricas e quantitativas para a elaboração das noções de qualidade; e pragmática, que considera que a qualidade da informação é essencialmente subjetiva e como tal, não é passível de ser operacionalizada.

Big Data		IoT	Qualidade da Informação	
3Vs	5Vs	Elementos	Categoria	Dimensão
Volume Variedade Velocidade	Volume Variedade Velocidade Valor Veracidade	Volume Variedade Velocidade Valor Veracidade Identidade Conectividade Manutenibilidade	Intrínseca	Credibilidade Livre de Erros Objetividade Reputação
			Representacional	Representação Concisa Representação Consistente Interpretabilidade Compreensividade
			Acessibilidade	Acessibilidade Facilidade de Manipulação
			Contextual	Quantidade de Dados Completude Relevância Tempestividade Valor Agregado
Fontes: Klein e Gorton (2015) e Marquesone (2017).		Fonte: Autores (2019)	Fontes: Bentancourt (2015) e Fagundes et al. (2017)	

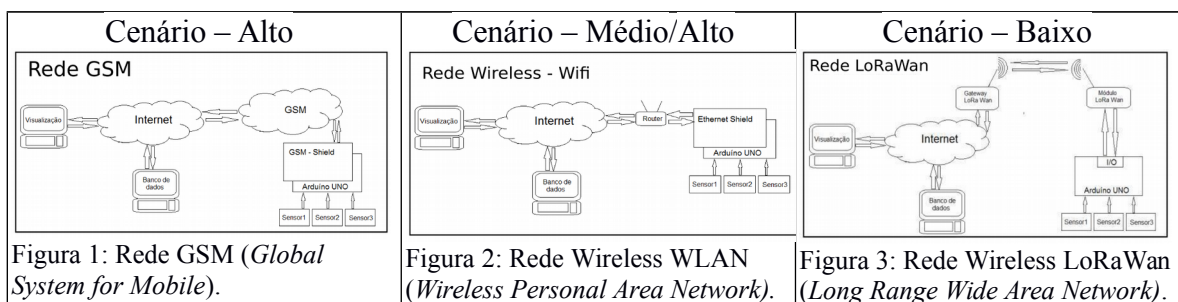
Quadro 1: Elementos essenciais para *Big Data* e *IoT* com as dimensões de qualidade da informação. Fonte: Autores (2019).

Também Bettancourt (2015), afirma que a qualidade de dados é tratada como sinônimo de qualidade da informação. Para os autores, a qualidade da informação está fortemente atrelada à qualidade dos dados, possuindo praticamente as mesmas dimensões de qualidade, com apenas algumas variantes. Ainda, conforme Assis e Moura (2011) e Santos *et al.* (2015), em muitas pesquisas dedicadas ao contexto dos sistemas de informação, a qualidade da informação é um termo substituível por qualidade de dados. Ainda, no quadro 1 são apresentados os elementos inerentes a cada área abordada. Neste trabalho foi possível

categorizar os elementos também para *IoT*, identificando prioridades para cada um dos pontos em comum ao *Big Data*, *IoT* em aderência a qualidade da informação.

3.2 INFRAESTRUTURA COLABORATIVA E INTEROPERÁVEL

O avanço da tecnologia tem permitido que incontáveis dados sejam trocados entre as pessoas por meio de diferentes ferramentas. Quando estas trocas apresentam um grande volume de dados variáveis, que podem ser analisado e processados em alta velocidade, toda essa carga de informação recebe o nome de *Big Data*. O gerenciamento de *Big Data* introduz desafios em termos de gerenciamento de dados, uma vez que os convencionais bancos de dados relacionais são ineficazes para o seu tratamento (Kumar *et al.*, 2014). Em especial o volume, variedade, velocidade, valor e veracidade são elementos fundamentais com a qualidade dos dados e da informação, são essenciais para obter sucesso na implementação do *Big Data*. Porém a comunicação de dados apresenta algumas limitações, poucas tecnologias de baixo custo, dentro de alguns cenários conforme as figuras 1, 2 e 3.



Fonte: Autores (2019).

Diferentes tecnologias de redes de comunicação oportunizam a implantação de redes sem fio em áreas de difícil acesso, a exemplo de matas, rios ou também áreas denominadas de alto risco. Embora o monitoramento realizado, ainda existem áreas sem monitoramento contínuo. Conforme Silva *et al.* (2016), diferentes cenários proporcionam que a *IoT*, possa consistir em uma rede de dispositivos físicos conectados. Estes dispositivos podem ser máquinas, partes de máquinas, sensores, atuadores, entre outros objetos com interfaces de comunicação, unidades de processamento e armazenamento que permitam que estes objetos mantenham comunicação. Com a inclusão dos dispositivos para *IoT*, apesar desta tecnologia inovadora trazer diversas facilidades na conexão de objetos, uma de suas necessidades é a existência de protocolos de comunicação para serem utilizados entre os dispositivos dentro de cada rede. Estes protocolos devem ser adequados dentro do contexto em que serão utilizados, como por exemplo, em redes cujos dispositivos, são alimentados por energia

proveniente de baterias, sendo necessário a escolha de um protocolo de comunicação que apresente utilize um baixo consumo de energia.

4 PROCEDIMENTOS METODOLÓGICOS

Visando identificar o efeito do fenômeno *Big Data* sobre o processo decisório, o método empregado é a pesquisa qualitativa através de levantamentos e apuração dos dados. Contudo, foram realizados levantamentos em campo para obtenção dos dados, juntamente com aquisição de novas bases cartográficas. Paralelo a esta etapa, com uso da plataforma *Hortonworks Sandbox HDP* (Hadoop Map-Reduce), conforme Bengfort e Kim (2016), proporcionando uma forma de integrar e distribuir o processamento envolvido com outras aplicações, a exemplo do i3Geo(2016), Geonetwork(2018), MapServer(2019), OSGEO com GeoServer(2019) e PostgreSQL/PostGIS(2019). Contudo, para o acesso e disponibilização de dados através de um portal, conforme Perin (2018) para gestão dos recursos hídricos, a bacia do Vale do Itajaí SC - Brasil e também dados de cidades ao entorno. Consultas aos dados foram realizadas em diferentes repositórios, a exemplo: Agência Nacional de Águas (ANA), Embrapa, Epagri, IBGE, INDE, Ministério de Meio Ambiente (MMA) e dos municípios da AMFRI (Associação dos Municípios da Foz do Rio Itajaí. Sendo possível obter resultados dinâmicos e incorporar novas classes para monitoramento do nível dos rios.

Para melhor interação dos decisores/gestores, novas interfaces de acesso aos dados e navegação foram avaliadas, a permitir melhor compreensão da informação com carga cognitiva mínima e maior precisão. Também a partir dos estudos de Aziyati *et al.* (2015), foram adotados indicadores para análise de *Big Data* tendo uma abordagem útil para resolução de muitos problemas. Neste contexto foram observadas as políticas do governo brasileiro principalmente da Infraestrutura Nacional de Dados Espaciais - (Inde, 2019).

4.1 COLETA DE DADOS E INFORMAÇÕES

Porém, a partir do estudo de caso foi possível a investigação empírica de um fenômeno recente, considerando o seu contexto real e suas muitas variáveis de análise (Hanton, 2001). Também Oliveira *et al.* (2009, p.36), indicando o estudo de caso para ter “uma visão detalhada de um fenômeno incluindo o seu contexto”, ideal quando se busca um melhor entendimento de um fenômeno desconhecido, possibilitando estudos posteriores para a construção de teorias. Contudo, as formas de obtenção de dados são discutidas por Marquesone (2017, p.36), e a partir da ideia de otimizar recursos através das melhores experiências,

foram elaboradas funções de pesquisa, avaliação e compartilhamento de informações, obtidas de estações telemétricas, sensores distribuídos e bases certificadas por diferentes agentes públicos.

Com isso, para o processamento de carga foram identificados vários conjuntos de dados em diferentes padrões, bancos de dados (postgresql/postgis, mysql/gis, oracle/spatial, msqlserver), shapefiles e dados de sistemas de monitoramento. Além de arquivos em papel e relatórios no formato (pdf, doc, odt e txt), documentos de planilhas (xls, ods e csv), dados de sensores (txt, csv, json e xml), imagens (satélites, mapas, fotos aéreas em papel, digital no formato png, jpg, tiff e geotiff) entre outros. Também diversos registros de evento dos logs de sensores, dispositivos móveis, servidores e de aplicações. Através da coleta autônoma com dispositivos *IoT*, foram implementados sensores e gateways, possibilitando a coleta e transmissão de atributos referente aos pontos de amostragem, contendo: localização, temperatura, umidade, nível do rio, dados de funcionamento de sensores (conectividade e consumo de energia).

Com serviços web foram coletados conjuntos de dados extraídos do monitoramento das estações telemétricas de acesso aberto (Ceops, 2019). Isto permitiu que todos os dados coletados fossem armazenados em repositório na nuvem e sua integração com outros sistemas legados com maior interoperabilidade. Na tabela 1 são apresentadas estimativas do volume na geração de registros provenientes de um período de coleta para o *Big Data*.

Tabela 1- Levantamento com Estimativas do Volume de Registros e Processamentos.



Descrição	Unidades	Atributos	Intervalo de Leitura (tempo/min)	Leitura de Registros						Volume Total (Registros)
				1 Hora	1 Dia	1 Mês	1 Ano	5 Anos	10 Anos	
Estações Telemétricas	11	5	15	4	96	2.880	34.560	172.800	1.728.000	19.008.000
Sensor Temperatura / Umidade	10	3	5	12	288	8.640	103.680	518.400	5.184.000	51.840.000
Sensor de Nível	30	5	10	6	144	4.320	51.840	259.200	2.592.000	77.760.000
Conectividade	10	3	5	12	288	8.640	103.680	518.400	5.184.000	51.840.000
Consumo de Energia	10	3	15	5	120	3.600	43.200	216.000	2.160.000	21.600.000
Manutenibilidade	10	4	30	2	48	1.440	17.280	86.400	864.000	8.640.000
Totais	81	23	80	41	984	29520	354240	1771200	17712000	222.048.000

Fonte: Autores (2019).

4.2 MODELAGEM DO SISTEMA

A modelagem dos artefatos e suas evoluções, possibilitaram também a realização da modelagem dos dados para viabilidade e padronização dos bancos de dados, abordando especificações e modelos disponibilizados por diferentes instituições públicas, conforme Guedes (2018), Lourenço *et al.* (2018) e Reggio (2018). Para o desenvolvimento, a figura 4 ilustra a estruturação da arquitetura dividida em módulos, componentes, subsistemas, processos e threads para implementações.

Conforme Vazquez *et al.* (2016), o diagramas a exemplo de “componentes” tem o objetivo de auxiliar a comunicação entre analistas e o *stakeholder* de forma que descreva o

cenário com as funcionalidades do sistema. Após a implementação para aquisição dos dados da área de estudo, a exemplo dos dados espaciais e meteorológicos. (Custódio, 2012). Ainda para cada estação, a interoperabilidade dos dados ocorreu dinamicamente no formato JSON Truong (2018). Assim, com armazenamento dos dados no servidor remoto, periodicamente os últimos registros de cada estação são inseridos em uma base de dados própria.

A redundância dos dados, visou prevenir o sistema de alerta em momentos de catástrofes sem precedentes, permitindo a disponibilização das informações de forma ágil em diferentes sistemas de monitoramento, garantindo a persistência dos dados, além de contribuir para diferentes estudos. Conforme a figura 5, ao se fazer acesso em dois sistemas de alerta a cada quinze minutos sequencialmente, ambos *web services* são processados para sincronizar objetos do tipo “medição”, sendo populados os atributos do objeto baseado nos dados de cada estação e posteriormente encaminhados ao repositório do *Big Data*.

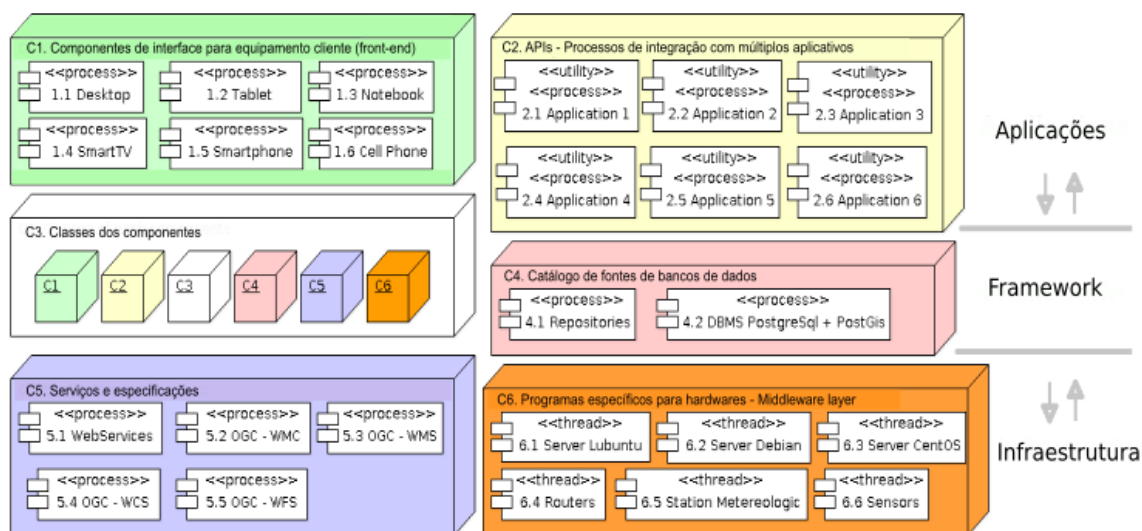


Figura 4: Diagrama de componentes do sistema – SADEPreaigeo.
Fonte: Autores (2019).



Ainda, conforme a figura 5 também é viabilizada na transmissão dos dados o código de identificação e a cidade origem da estação, além de dados mais analíticos que são também adicionados na base dados. a exemplo da data da medição, valor do nível do rio em metros, valor de precipitação em milímetros, dois campos booleanos (retornam se o “nível da chuva” e do “rio” estão ativos ou não) e por último o estado da estação.

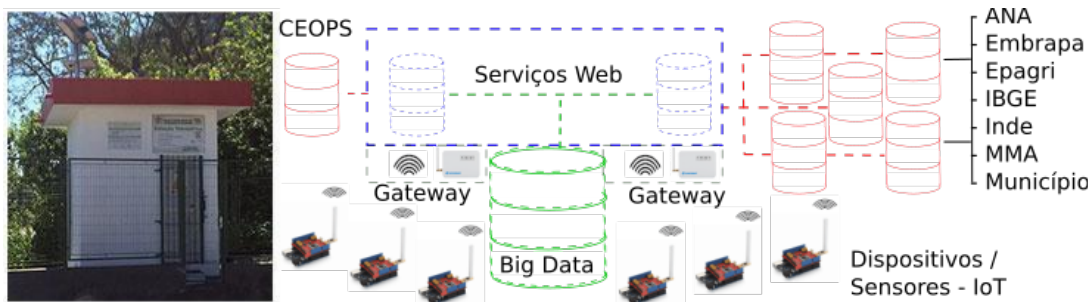


Figura 5: Arquitetura de integração.
Fonte: Autores (2019).

Podendo também ser adotadas análises de imagens utilizando os atributos espectrais, espaciais e hierárquicos. (Körting, 2018). Possibilitando investigações minuciosas sobre dados coletados pelas diferentes agências públicas e também privadas.

5 ANÁLISE E DISCUSSÃO DOS DADOS

Após as etapas anteriores e da coleta dos dados, foram realizadas as primeiras análises dos dados e das informações, alcançando os primeiros resultados com a estruturação básica do *Big Data* proposto. As questões relacionadas a catalogação dos dados, teve como propósito auxiliar na classificação da informação e procurou identificar relações entre os fenômenos, os quais serão melhor discutidos em trabalhos futuros.

Porém, a partir das observações sobre diferentes fontes de conjuntos de dados foi possível realizar leituras sobre as ocorrências e movimentos nestes conjuntos, e posteriormente adequação dos primeiros algoritmos para ativação e sincronização dos mesmos. Considerando a necessidade de certificação de conjuntos dependentes e independentes, ou seja, o tratamento de vetores de dados, sendo $S = \{db_1, db_2, \dots, db_n\}$ no qual foi gerado um novo espaço vetorial, a cada vetor escrito como uma combinação linear de S . Contudo, embora possa haver mais de uma maneira de expressar um vetor a partir de um conjunto gerador, foram geradas regras observando um conjunto não vazio de vetores, conforme a equação: $\{k_1db_1 + k_2db_2 + \dots, k_ndb_n = 0\}$. Havendo pelo menos uma solução para $k_i=0$ e sucessivamente, ainda se esta é a única solução, então o conjunto S é denominado linearmente independente, caso contrário é denominado linearmente dependente.

Com as regras definidas, permitiu gerar os parâmetros na plataforma Hadoop-MapReduce, sendo implementados no estudo de caso com a representação das informações no formato de gráficos. Com esta função o decisor pode visualizar os gráficos com atualização das últimas vinte e quatro horas. Os gráficos são gerados por uma API chamada *D3.js*, sendo uma biblioteca para produzir visualizações de dados dinâmicos de forma

interativa nos navegadores (*browsers*). Mas, juntamente com os dados disponíveis no *dashboard*, foi importante referenciar os detalhes de uma única estação e comparar gráficos de estações distintas da mesma cidade. Também procedimentos para emissão de alertas foram preparados, a exemplo de (e-mail e SMS), contendo detalhes dos níveis de rios e tendo como foco principal ajudar na divulgação da informação aos moradores e da população em torno de perímetros considerados de alto risco.

Através do presente trabalho foi possível detectar a concepção de uma estrutura elementar para implementação de um *Big Data*, visando ampliar ações e a gestão do conhecimento. É fundamental que os decisores/gestores saibam lidar com as características das informações de hoje, no contexto do *Big Data* – volume, variedade, velocidade, valor e veracidade aliada a dimensões para maior qualidade da informação. Porém, para isso, é necessário que reflitam e compreendam como essas características influenciam no seu processo de decisão e, consequentemente, o processo de decisão da organização.

Neste contexto, o presente trabalho propôs a realização da revisão da abordagem para intensificar a coleta e a integração com dados geoespaciais e verificar o efeito dos fatores 3V's (volume, variedade, velocidade). Também devido a maior parte dos processos de alguma forma utilizam a conexão com a internet alimentando o *Big Data*, se pode confirmar que suas ações (valor, veracidade) dos dados, contribuem para o 5V's no processo de tomada de decisão. Com isso, através de estudos foram confirmadas ao longo da pesquisa qualitativa formas de se atingir os objetivos específicos: (i) levantamento de soluções de Infraestrutura de hardware e software para implementar repositório para grandes volumes de dados; (ii) detecção de soluções integradas para implementar mecanismos autômatos; (iii) revisão de métodos supervisionados e não supervisionados para análise dos dados; e (iv) fundamentação teórica para o desenvolvimento de mecanismos a propagação de acesso a informação distribuída.

6 CONSIDERAÇÕES FINAIS

Foram apresentados neste artigo importantes aspectos conceituais para contribuir no ciclo de gerenciamento do conhecimento distribuído, apoiado pela metodologia de engenharia do conhecimento CommonKADS para apontar os elementos principais de uma estrutura para implementação de um *Big Data*. Sendo assim, ao final, o objetivo principal foi alcançado, tendo como resultado a organização das diretrizes de implementação. Além de indicar e priorizar as variáveis (volume, variedade, velocidade, valor e veracidade) para o processo decisório. As variáveis que mais influenciam o processo de tomada de decisão são

“veracidade” e “valor” das informações analisadas. Após, as três variáveis que menos impactam, respeitando a ordem final, são variedade, velocidade e volume de informações. Como contribuição através deste trabalho, fica a oportunidade do mesmo ser aproveitado no momento de formulação das prioridades, bem como, para reflexão a respeito do uso e do tratamento da informação no seu processo de decisão, minimizando assim, os riscos em uma tomada de decisão com ausência dos critérios de integração e também sem fundamentação a partir dos dados.

A relevância científica do presente trabalho, é detectada na organização dos dados e proporciona aos usuários do *Big Data* acesso aos dados estatísticos para melhor entendimento de fenômenos naturais. Também vale destacar a relevância tecnológica, seguindo padrões conforme apresentado na figura 4, que por sua vez, proporciona arquitetura do portal integrado ao sistema SAD-PREAIgeo (<http://sadpreaigeo.org>). Também a relevância social, promove ações e mecanismos para atender carências de acesso a informações de domínio público de pequena, média e alta complexidade. Compartilhando processos para uma análise quantitativa do padrão espacial relacionado ao uso da terra através de dados DEM (*Digital Elevation Model*), modelo digital de elevação. Com o crescimento de grandes volumes de dados, o trabalho de refino e análise através do *Big Data Analytics* se faz indispensável e muito estratégico para tomada de decisão. Sendo assim, no mundo dos dados de hoje, às vezes não é preciso de grandes volumes de dados, mas sim de uma pequena dose de coerência com os dados corretos, elevando com precisão a qualidade das informações.

AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- Assis, J.; Moura, A. (2011). A Qualidade da Informação na Web: uma abordagem semiótica; La Calidad de la Información en la Web. *Revista Informação & Informação*, 16, 96-117.
- Aziyati, Y., Norashidah, M. D., Salman Y. (2015). Big data analytics for Flood Information Management in Kelantan, Malaysia. *IEEE Student Conference Research and Development*.
- Barbieri, C. (2013). *Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)*. Belo Horizonte: Fumsoft.
- Berman, J.J. (2013). *Principles of big data: preparing, sharing, and analyzing*. Newnes Editor.

- Bengfort, B., Kim, J. (2016). *Analítica de dados com Hadoop: Uma introdução para cientistas de dados*. São Paulo, Novatec.
- Bentancourt, M.P.S. (2015). Servqual como Instrumento de Gestão da Qualidade da Informação em Ambiente EaD. *Dissertação*. Universidade Federal de Santa Catarina.
- Bertone, P., & Marinho, C. (2013). Gestão de Riscos e Resposta a Desastres Naturais: A visão do Planejamento. *VI Congresso de Gestão Pública - CONSAD*, 1-24.
- Britto, S.M., Almeida J.R.J. (2006). Qualidade de Dados para Data Warehouse: Roteiro de Implementação. *Congresso Internacional de Gestão da Tecnologia e S.I.*. 255-272.
- Buffon, E.A.M., Paz, O.L.S. (2017). Avaliação das áreas com risco de inundações em Curitiba, Paraná Brasil, a partir do cruzamento dos mapeamentos de cobertura da terra e de vulnerabilidade social. *XVIII Simpósio Brasileiro de Sensoriamento Remoto*, 5552-5559.
- Ceops, Centro de Operações do Sistema de Alertas. (2019) *[Sistema OnLine]*. Recuperado em 13 junho 2019, de <http://ceops.furb.br/>
- Cunha, M.A.V.C. (2018). O uso da tecnologia da informação e comunicação para caracterizar os municípios: quem são e o que precisam para desenvolver ações de governo eletrônico e smart city. *Revista Administração Pública*, 52, 4, 630-649.
- Custodio, D.M. (2012) Mesomóvel: Uma Estação Meteorológica Móvel de Superfície para a Observação de Fenômenos Convectivos Locais. *Dissertação*. Univ. Federal de Santa Maria.
- Dama - Data Management Association International. (2012). *Guide (DAMA-DMBOK) Portuguese Edition*. Recuperado em 13 junho 2019, de <http://www.dama.org.br/conhecimento/>
- Fagundes, V.B., Bentancourt, S.M.P., Bastos, L.C. (2017). Qualidade da Informação na Revisão Sistemática de Literatura. *VII Congresso Internacional de Conhecimento e Inovação*, 39.
- Guedes, G.T.A. (2018). *UML 2: Uma abordagem prática*. (3a ed.) São Paulo, Novatec.
- Hai, S., Cheng, W., Bo, R., (2008). Method of DEM Data's Processing in Flood Simulation System. *Computer Science and Software Engineering, IEEE International Conference*, 50-53.
- Hanton, Howard. (2001). *Álgebra linear com aplicações*. Porto Alegre, Bookman.
- Inde - Infraestrutura Nacional de Dados Espaciais. (2019) *Perfil de Metadados Geoespaciais do Brasil*. Recuperado em 13 junho 2019, de <https://inde.gov.br/>
- Klein, J., Gorton, I. (2015). Runtime performance challenges in big data systems. *Proceedings Workshop on Challenges in Performance Methods for Software Development*. 17-22.
- Kon, F., & Santana, E. F. (2016). *Cidades Inteligentes: Conceitos, plataformas e desafios*. Congresso da Sociedade Brasileira de Computação (pp. 2 - 49). Porto Alegre: PUCRS.
- Kumar, R.I., Parashar, B.B., Gupta, S., Sharma, Y., Gupta, N. (2014). Apache hadoop, nosql and newsql solutions of big data. *International Journal of Advance Foundation and Research in Science & Engineering*, 1(6), 28-36.
- Körting, T.S. (2018). Análise de imagens baseada em objetos geográfico com GEOBIA. *INPE - SELPER-2018*. Recuperado em 13 junho 2019, de <https://moodle.dpi.inpe.br/>
- Lemos, A. (2013). De que forma as novas tecnologias como a computação em nuvem, o Big Data e a IoT-podem melhorar a condição de vida nos espaços urbanos? *GVexecutivo*, 12(2), 46-49.
- Lourenço, G.C.U., Sartori, R., Menegassi, C. H. M., Oliveira, & M., Tenório, N. (2018). Os Artefatos de Software Enquanto Produtos do Conhecimento: Uma Investigação na Indústria de Software. *VIII Congresso Internacional de Conhecimento e Inovação*, 277.
- Malaquias, F.S.S., Teixeira, L.M.S., Santiago, R.B.S., Parente, L.L., Júnior, L.G.F. (2017). Pastagem.org: dados e informações para uma compreensão ambiental, econômica e social sobre a última fronteira agrícola do Brasil. *XVIII Simpósio Brasileiro de Sensoriamento Remoto*, 1392-1399.

- Marquesone, R. (2017). *Big Data - Técnicas e tecnologias para extração de valor dos dados*. São Paulo. Casa do Código.
- Meireles, L.V. (2018). Projeto de uma rede de internet das coisas para monitoramento e alerta de emergência em áreas de risco. *Dissertação*. Universidade Federal de Ouro Preto.
- Melhado, A.R., Pasini, R. (2018, Jun). Engenharia do futuro passa por sustentabilidade e inovação. Recuperado em 13 junho 2019, de <https://cimentoitambe.com.br/engenharia-do-futuro-passa-por-sustentabilidade-e-inovacao/>
- Montori, F., Bedogni, L., Bononi, L. (2018, April). A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. *Internet of Things Journal*, 5-2, pp. 592-605, Recuperado em 18 junho, 2019, de <https://ieeexplore.ieee.org/document/7961139>
- Oliveira, M., Maçada, A.C., Goldoni, V. (2009). Forças e fraquezas na aplicação do estudo de caso na área de sistemas de informação. *Revista de Gestão USP*, São Paulo, 16(1), 33-49.
- Palandi, F., Gerges, N.R., Costa, E. (2018). Sustentabilidade Aplicada: O Papel do designer no desenvolvimento urbano baseado em conhecimento. *VIII Congresso Internacional de Conhecimento e Inovação*, 290.
- Perin, A., Soriano-Sierra, J. (2018). Gestão Integrada de Recursos Hídricos: Uma Revisão Integrativa. *VIII Congresso Internacional de Conhecimento e Inovação*. 263.
- Reggio, G.A. (2018). UML-based Proposal for IoT System Requirements Specification. *IEEE International Workshop on Modelling in Software Engineering*, 9-16.
- Ramos, A. (2015). *Infraestrutura Big Data com OpenSource*. Rio de Janeiro, Ciência Moderna.
- Santos, W.M.J., Costa, V.C. (2015). Geoinformação: disponibilização e qualidade de dados apresentados em ambiente de sistema e informação geográfica na internet (SIGWEB). *XVII Simpósio Brasileiro de Sensoriamento Remoto*, 3470-3478.
- Schreiber, August Th. (2000). *Knowledge engineering and management: the CommonKADS methodology*. MIT press.
- Schreiner, G.A. (2016). SQLTOKEYNOSQL: Uma camada para mapeamento de esquemas relacionais e de operações SQL para Bancos de Dados NOSQL baseados em chaves de acesso. *Dissertação* - Universidade Federal de Santa Catarina.
- Shen, H. (2011). Land use spatial pattern characteristics along the terrain gradient in Yellow River Basin in west Henan province, China. *IEEE - International Conference Geoinformatics*, 1-5.
- Silva L.A., Vieira F.M.F., Moraes A.F. (2016). Interoperability of Systems and Monitoring for Alert Itajaí River Basin-SaDPrai. *Advances in Intelligent Systems and Computing*. 444, 179-185.
- Simon B., Simon C., Jean R. (2018) CREATE: Multimodal Dataset for Unsupervised Learning and Generative Modeling of Sensory Data from a Mobile Robot, Recuperado em 20 junho, 2019, de IEEE Dataport.
- Sun, Y., & Du, Y. (2017) Big data and sustainable cities: applications of new and emerging forms of geospatial data in urban studies. *Open geospatial data, software standards*. 2 -24.
- Taurion, C. (2013). *Big Data*. Rio de Janeiro: Brasport.
- Truong, H. (2018) Integrated Analytics for IoT Predictive Maintenance using IoT Big Data Cloud Systems. *IEEE International Conference on Industrial Internet (ICII)*. 109-118.
- Vazquez. C.E., Simões, G.S. (2016). *Engenharia de Requisitos: software orientado a negócios*. Rio de Janeiro, Brasport.
- Zanella, A., Vangelista, L. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, 1, 22-32.