

Integração, Relacionamento e Representação de Dados em Cidades Inteligentes: Uma Revisão de Literatura

Larysse Silva, José Alex Lima, Nélio Cacho, Eiji Adachi,
Frederico Lopes, Everton Cavalcante

Universidade Federal do Rio Grande do Norte (UFRN)
Natal-RN, Brasil

larysse.savanna@ufrn.edu.br, j.alex.medeiros@gmail.com,
neliocacho@dimap.ufrn.br, {eijiadachi, fred}@imd.ufrn.br
everton@dimap.ufrn.br

Resumo. *Uma característica notável de cidades inteligentes é o aumento da quantidade de dados gerados produzidos pelos mais diversos dispositivos e sistemas computacionais, ampliando assim os desafios do desenvolvimento de software que envolva a integração de grandes volumes de dados. Nesse contexto, este artigo apresenta uma revisão de literatura a fim de identificar as principais estratégias utilizadas no desenvolvimento de soluções de integração, relacionamento e representação de dados em cidades inteligentes. Este estudo selecionou e analisou de forma sistemática onze artigos publicados entre os anos de 2015 e 2017. Os resultados obtidos evidenciam lacunas com relação a soluções para a integração contínua de fontes de dados heterogêneas a fim de dar suporte ao desenvolvimento de aplicações e tomada de decisão.*

Abstract. *A notable characteristic of smart cities is the increase in the amount of available data generated by several devices and computational systems, thus augmenting the challenges related to the development of software that involves the integration of larges volumes of data. In this context, this paper presents a literature review aimed to identify the main strategies used in the development of solutions for data integration, relationship, and representation in smart cities. This study systematically selected and analyzed eleven studies published from 2015 to 2017. The achieved results reveal gaps regarding solutions for the continuous integration of heterogeneous data sources towards supporting application development and decision-making.*

1. Introdução

A pesquisa e desenvolvimento de soluções em cidades inteligentes são iniciativas mundiais que levam a explorar melhor os recursos de uma cidade a fim de melhorar a qualidade dos serviços para os cidadãos. Os últimos avanços da Tecnologia da Informação (TI) têm proporcionado uma diversidade de tecnologias de *hardware* e *software* que resultam na produção de uma quantidade massiva de dados. Um relatório publicado em 2014 pela *International Data Corporation* (IDC) comprova esse fato ao informar que o volume total de dados gerados e transmitidos no mundo em 2013 foi de 4.4 ZB e a previsão para 2020 é que esse número chegue a ser dez vezes maior [Turner et al. 2014].

De um modo geral, inclusive no contexto de cidades inteligentes, a produção, coleta e disponibilização de dados geralmente são realizadas de forma distribuída, isolada

e sem um padrão definido. Isso ocorre devido à grande heterogeneidade de *stakeholders* envolvidos nas soluções para cidades inteligentes, tais como cidadãos, governo, indústria, academia etc., que geram e consomem dados e informações relacionadas a seus próprios interesses. Outro fator que contribui para essa diversidade é a ausência de políticas relacionadas ao compartilhamento e disponibilização desses dados, que geralmente possuem formatos definidos de acordo com a facilidade ou a capacidade técnica dos seus respectivos produtores [Souza et al. 2017]. Outro aspecto importante diz respeito à natureza desses dados, que são diversos, podem não estar devidamente estruturados e conter ruídos ou imprecisões que afetam de forma negativa a sua utilização por aplicações e usuários finais.

Devido ao aumento exponencial da quantidade de dados gerados por empresas, pessoas e dispositivos computacionais, o termo *Big Data* vem sendo utilizado para descrever enormes conjuntos de dados que, diferente dos dados tradicionais, geralmente incluem dados não estruturados que software de processamento de dados tradicional ainda não consegue lidar de forma adequada [Chen et al. 2014]. Tais dados são adquiridos a partir de diversos tipos de atividades, a exemplo da captação de medidas por sensores, publicações de usuários em redes sociais, envio de *e-mails* e transações bancárias ou mesmo dados extraídos de arquivos de *log* de servidores Web.

O principal objetivo de uma solução de *Big Data* é oferecer uma abordagem abrangente para o tratamento de volumes de dados caóticos para tornar as aplicações mais eficientes e precisas, transformando dados brutos em informação de valor agregado e conhecimento para os usuários. Além disso, o paradigma de *Big Data* também traz novas oportunidades para descobrir novos valores, auxiliar na obtenção e compreensão mais aprofundada dos valores ocultos, bem como integrar e relacionar dados de diferentes fontes, possibilitando o gerenciamento dos conjuntos de dados de forma efetiva.

Tão importante quanto gerenciar grandes volumes de dados é a sua integração para que seja possível extrair informações relevantes e transformá-las em conhecimentos práticos. Entretanto, atualmente é comum que a produção de dados ocorra de forma não-padronizada, tendo em vista que os sistemas são desenvolvidos de acordo com as necessidades e interesses dos desenvolvedores ou proprietários do sistema, que muitas vezes não levam em consideração outros atores que possam estar interessados nos mesmos dados. Outro fato é que o formato dos dados gerados é definido de acordo com a facilidade ou a capacidade técnica dos desenvolvedores, fazendo com que a integração dos dados seja um grande desafio para aplicações que precisam realizar consultas em diversas fontes de dados independentes e heterogêneas.

Nesse contexto, este trabalho tem como objetivo realizar uma revisão de literatura a fim de investigar as principais estratégias e ferramentas que permitem a integração, relacionamento e representação de grandes volumes de dados no contexto de cidades inteligentes. Para esse fim, foi adotada uma metodologia bem definida que permitiu a busca, seleção e análise sistemática dos estudos disponíveis na literatura [Kitchenham e Charters 2007]. A ideia é que os resultados obtidos permitam não só ter uma visão geral acerca das estratégias atualmente existentes para a integração dados gerados por diversos sistemas no contexto de cidades inteligentes, mas também elencar questões importantes a serem endereçadas em pesquisa e desenvolvimento no futuro.

O restante deste artigo está estruturado da seguinte forma. A Seção 2 discute brevemente alguns trabalhos relacionados a esta pesquisa. A Seção 3 apresenta a metodologia adotada neste trabalho em termos das questões de pesquisa a serem respondidas e as estratégias para busca e seleção dos estudos. A Seção 4 descreve como os estudos relevantes foram selecionados. A Seção 5 provê uma síntese resultante da análise dos estudos selecionados como respostas às questões de pesquisa estabelecidas. A Seção 6 levanta possíveis ameaças à validade da revisão de literatura realizada. Por fim, a Seção 7 apresenta algumas conclusões a partir dos resultados encontrados e direções para trabalhos futuros.

2. Trabalhos Relacionados

É possível encontrar na literatura trabalhos relacionados ao processamento de dados em cidades inteligentes, porém há uma escassez de estudos ligados diretamente a integração, relacionamento e representação de dados em larga escala. A seguir são apresentados alguns trabalhos relacionados encontrados na literatura.

Macêdo *et al.* (2017) fizeram um estudo comparativo sobre ferramentas existentes para publicação e gerenciamento de dados abertos em cidades inteligentes. Como resultado, os autores concluíram que as ferramentas apresentadas não foram capazes de atingir adequadamente critérios como facilidade de uso por usuários sem conhecimento técnico especializado, possibilidade de mapeamento objeto-relacional e qualidade dos dados em termos de completude, confiabilidade e precisão. Por fim, os autores propuseram como trabalho futuro a implementação de uma nova ferramenta visando diminuir a complexidade e os custos da publicação e do gerenciamento de dados abertos.

Chen *et al.* (2015) realizaram uma revisão de literatura sobre grandes volumes de dados no contexto da Internet das Coisas (IoT). Os autores forneceram uma maneira sistemática de analisar a mineração de dados em visão de conhecimento, visão técnica e visão de aplicação, incluindo classificação, agrupamento, análise de associação, análise de série temporal e análise isolada, incluindo os casos mais recentes de aplicações na pesquisa. Os autores analisaram ainda os algoritmos mais recentes de análise de dados a fim de aplicá-los a soluções de *Big Data* e discutiram desafios de pesquisa futura.

3. Metodologia

Uma revisão de literatura permite analisar estudos disponíveis em um determinado domínio e permitem responder a questões de pesquisa sobre o atual estado da arte da literatura ou estado da prática. Contudo, a fim de conferir valor científico a tal revisão, ela precisa ser feita seguindo um procedimento rigoroso e sistemático com vistas a minimizar possível viés e tornar o processo reproduzível por outros pesquisadores. Para isso, é necessário estabelecer um protocolo bem definido com questões de pesquisa e critérios explícitos para avaliar e selecionar os estudos disponíveis na literatura, protocolo esse que precisa ser seguido de forma estrita ao longo de todo o processo [Kitchenham e Charters 2007].

A revisão de literatura apresentada neste trabalho foi conduzida em conformidade com metodologias bem estabelecidas na literatura [Biolchini e Travassos 2005, Kitchenham e Charters 2007]. Como ilustra a Figura 1, as etapas realizadas foram basicamente

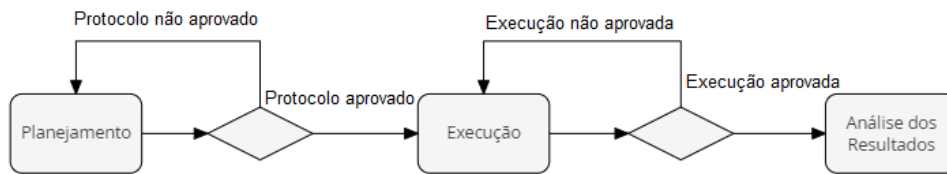


Figura 1. Fases do processo de revisão de literatura

três: (i) *planejamento*, na qual foi elaborado um protocolo definindo as questões de pesquisa a serem respondidas, a estratégia de busca adotada, os critérios utilizados para a seleção dos estudos e os métodos para extração e síntese de dados; (ii) *execução*, na qual os estudos foram identificados, selecionados e avaliados de acordo com o protocolo estabelecido, e; (iii) *análise dos resultados*, na qual as informações extraídas dos estudos selecionados foram agregadas considerando as questões de pesquisa, bem como feita uma análise e discussão acerca dos resultados encontrados.

Questões de pesquisa. O objetivo principal desta revisão de literatura foi identificar estudos relevantes que apresentassem contribuições na elaboração de estratégias para integração, relacionamento e representação de grandes volumes de dados de diferentes fontes no contexto de cidades inteligentes. Com esse objetivo em vista, foram definidas as seguintes questões de pesquisa (QPs):

- QP1: Quais as principais estratégias adotadas para o desenvolvimento de sistemas de integração, relacionamento e representação de dados em cidades inteligentes?
- QP2: Quais as principais ferramentas utilizadas na integração, relacionamento e representação de dados em cidades inteligentes?
- QP3: Quais os principais desafios encontrados no desenvolvimento de sistemas de integração, relacionamento e representação de dados?

Estratégia de busca. Para recuperar os estudos relacionados ao objetivo central deste trabalho, foi utilizado um processo automatizado de busca sobre três bases eletrônicas de publicações científicas, a saber, ACM Digital Library¹, IEEEExplore² e Google Scholar³. Com base nas questões de pesquisa, foi elaborada a seguinte *string* de busca:

```

(data integration E data representation E data relationship) E
(smart city OU smart cities)
  
```

Critérios de seleção. Critérios de seleção foram utilizados para avaliar cada um dos estudos recuperados a fim de incluir apenas estudos de fossem de fato relevantes para responder às QPs e excluir aqueles que não contribuíssem para respondê-las.

Foram considerados os seguintes critérios de inclusão (CIs):

- CI1: O tema do estudo está relacionado a integração, relacionamento e representação de dados em cidades inteligentes.

¹<http://dl.acm.org>

²<http://ieeexplore.ieee.org>

³<http://scholar.google.com>

CI2: O estudo apresenta uma estratégia para a integração, relacionamento e representação de no mínimo dois tipos de dados diferentes.

Foram estabelecidos ainda os seguintes critérios de exclusão (CEs):

CE1: O estudo não está relacionado ao tema de integração de dados.

CE2: O estudo descreve uma abordagem sem detalhes claros e suficientes sobre como realizar a integração, relacionamento e representação de dados.

CE3: O estudo não foi publicado entre janeiro de 2015 e dezembro de 2017.

CE4: O estudo é uma duplicação de outro resultado já recuperado ou um trabalho mais recente em comparação a outro anteriormente publicado.

CE5: O resumo e/ou o texto completo do estudo não está disponível.

CE6: O estudo não foi publicado em Inglês, que é o idioma mais comumente utilizado em publicações científicas.

Nesta revisão de literatura, um estudo foi considerado relevante se este não atendeu a nenhum dos CEs e atendeu a pelo menos um IC.

4. Processo de Seleção

Para recuperação dos estudos a partir das bases eletrônicas de publicação, a *string* de busca foi adaptada a fim de torná-la compatível com as especificidades dos mecanismos de busca de cada uma das bases eletrônicas. Feito isso, o procedimento de busca automatizada foi realizado sobre cada base eletrônica de acordo com a *string* de busca adaptada. A busca automatizada foi limitada apenas aos campos de título, resumo e palavras-chave.

Após recuperar os estudos a partir das bases eletrônicas, o processo de seleção foi conduzido em três etapas. Na Etapa 1, o título e do resumo do estudo foram lidos e analisados em conformidade com os critérios de seleção estabelecidos no protocolo. Na Etapa 2, os estudos filtrados na etapa anterior tiveram as seções de introdução e conclusão lidas e analisadas para verificar se o tema abordado no estudo era realmente compatível com o tópico de interesse deste estudo. Por fim, na Etapa 3, os estudos filtrados na etapa anterior foram lidos por completo, seguindo-se atividades de extração de dados para apoiar a elaboração de respostas às questões de pesquisa anteriormente estabelecidas.

Após a aplicação da *string* de busca nas bases eletrônicas, foram encontradas 290 publicações científicas no total, sendo 264 da IEEEXplore, 25 da ACM Digital Library e 1 do Google Scholar. A Tabela 1 apresenta uma evolução da seleção dos estudos na execução das três etapas do processo de seleção. A primeira coluna apresenta a quantidade de artigos identificados na Etapa 1, a segunda coluna exibe os artigos selecionados na Etapa 2 e, por fim, a quantidade de artigos extraídos na Etapa 3.

Tabela 1. Resultados das etapas do processo de seleção dos artigos

Base eletrônica	Etapa 1	Etapa 2	Etapa 3
IEEEXplore	264	38	8
ACM Digital Library	25	4	3
Google Scholar	1	0	0
Total	290	42	11

5. Resultados

Esta seção sumariza os resultados da revisão de literatura realizada considerando as questões de pesquisa e os dados extraídos/sintetizados a partir dos estudos selecionados. As Seções 5.1 a 5.3 apresentam respostas a cada uma das QPs inicialmente definidas no protocolo em termos de estratégias, ferramentas e desafios para a integração, relacionamento e representação de dados em cidades inteligentes.

5.1. Estratégias (QP1)

Dentre as principais estratégias identificadas nesta pesquisa, podem ser citadas:

Souza *et al.* (2017) desenvolveram um *middleware* de dados chamado *Smart Geo Layers* baseado em dados geoespaciais que visa unificar os dados fornecidos por diversas fontes em ambientes de cidades inteligentes, permitindo que usuários de diversas organizações possam compartilhar e consumir dados. Os autores desenvolveram um modelo unificado de dados para facilitar a integração de dados de diferentes tipos e formatos. A arquitetura da solução proposta pelos autores é composta por elementos que dispõem de *APIs RESTfull* para consumir e enviar dados para o *middleware*, um *context broker* que é de um componente intermediário para armazenar camadas registradas no *Smart Geo Layers*, um componente de segurança para realizar autenticação de usuários a fim de proteger dados privados e também foi desenvolvido um componente responsável por traduz as operações fornecidas pelas *APIs* para as linguagens de consulta específicas do sistema de banco de dados (NoSQL ou geoespacial). O banco de dados *NoSQL* armazena atributos não-geográficos das entidades armazenadas no *Smart Geo Layers*, enquanto o banco de dados geoespacial armazena valores de atributos geográficos. Os autores informam que essa separação dos dados visa garantir a escalabilidade, pois as consultas que usam funções geográficas são mais complexas e demoradas do que as consultas com atributos apenas textuais.

Chakraborty *et al.* (2017) destacam o uso da abordagem *Extract-Transform-Load* (ETL, do inglês, Extração, Transformação e Carregamento) para integração de dados. O ETL é um processo de armazenamento de dados que extrai dados de fontes externas, transforma-os em necessidades operacionais, que podem incluir verificações de qualidade e carrega-os no banco de dados de destino. Esse processo é dividido em três fases:

- Extração: nessa primeira fase ocorre o processo de extração dos dados das fontes. Normalmente esses dados são extraídos em formato simples como csv, xls e txt ou por meio de um cliente *RESTfull*.
- Transformação: essa fase compreende a limpeza dos dados, ou seja, remoção de dados duplicados, verificação de violação de integridade, classificação e agrupamento de dados, etc.
- Carregamento: essa fase envolve a persistência dos dados na base consolidada.

Com base no estudo elaborado pelo autor sobre a abordagem ETL, é possível identificar que essa estratégia possui diversas limitações como, por exemplo, a impossibilidade de automatizar as soluções que fazem uso do ETL, havendo a necessidade da intervenção manual principalmente na etapa de transformação dos dados. Outra lacuna encontrada nessa estratégia é a capacidade limitada para extrair dados de diferentes fontes ao mesmo tempo, pois embora na maioria das ferramentas de ETL a consulta possa ser

criada e utilizada, o que é muito semelhante à consulta SQL, só é possível extrair dados de fonte de dados única.

You *et al.* (2017) fizeram uso da abordagem *Informed Design Platform* (IDP) para realizar a coleta, armazenando, limpeza, análise, integração, mineração e visualização de dados. Essa abordagem é composta por cinco componentes, a saber, (i) *objetos*, (ii) um coletor para reunir dados dos objetos, (iii) uma plataforma central de gerenciamento de dados e permitir uma integração de sistemas fracamente acoplada, (iv) de serviços distribuídos modularizados e reutilizáveis para processar dados de várias fontes e (v) um orquestrador de serviços.

5.2. Ferramentas (QP2)

Dentre as principais ferramentas utilizadas nos trabalhos encontrados, pode-se destacar:

O Clover ETL, que é uma plataforma ETL de integração de dados baseada em Java para rápido desenvolvimento e automação de transformações de dados, limpeza de dados, migração de dados e distribuição de dados em aplicações, bancos de dados e armazenamento em nuvem.

O Talend, que é uma ferramenta de integração de dados de código aberto desenvolvido pela Talend e projetado para combinar, converter e atualizar dados em vários locais nos negócios. O Talend Open Studio para Integração de Dados funciona como um gerador de código, produzindo scripts de transformação de dados e programas subjacentes em Java.

O Pentaho Data Integration (PDI) é o componente do Pentaho responsável pelos processos de ETL e pode ser usado para migrar dados entre aplicativos ou bancos de dados, exportar dados de bancos de dados para arquivos simples, carregar dados maciçamente em bancos de dados, limpeza de dados e integração de dados de aplicações. Todo processo no PDI é criado com uma ferramenta gráfica onde o usuário especifica o que fazer sem escrever linhas de código, tornando-o orientado a metadados. Apache Hadoop é uma ferramenta para processamento distribuído de grandes conjuntos de dados em computadores usando modelos de programação simples.

A ferramenta GUIDES é apresentada por Balasubramani *et al.* (2017) como uma nova estrutura de conversão e gerenciamento de dados para sistemas de infraestrutura urbana subterrânea que permite que administradores municipais, trabalhadores e contratados, juntamente com o público em geral e outros usuários, consultem dados digitalizados e integrados para tomar decisões mais inteligentes.

O FIWARE foi utilizado por Souza *et al.* (2017) e é uma plataforma genérica e extensível capaz de lidar com os requisitos essenciais em cidades inteligentes. Essa plataforma oferece componentes que fornecem um ambiente para suportar o desenvolvimento de aplicações que precisam integrar dados heterogêneos, aplicações lógicas e elementos da interface do usuário baseados na Web, entre outras funcionalidades como realizar a autenticação e o gerenciamento de identidades e credenciais de usuários, organizações e aplicações, implementação de interfaces que permitem acessar, editar, visualizar e analisar informações de multicamadas que podem ser representadas por dados geoespaciais.

Sobre as limitações das ferramentas mencionadas, é possível citar que as ferramentas CloverETL e Pentaho não fornecem suporte web semântico para a visualização

do modelo de dados, isso se torna uma limitação tendo em vista que a possibilidade de o usuário selecionar as classes e propriedades adequadas para fornecer relacionamentos entre elas através da visualização é uma parte essencial de um ETL semântico.

5.3. Desafios (QP3)

Os principais desafios encontrados nas soluções de integração, relacionamento e representação de dados são em relação a heterogeneidade, inconsistência, instabilidade e atualização frequente dos dados. A heterogeneidade geralmente é causada por diversos fatores, como dados autônomos, diferenças nas fontes de dados, técnicas de coleta e armazenamento de dados, etc. A inconsistência dos dados é causada principalmente por imperfeições que acabam afetando na qualidade desses dados. Um exemplo disso é quando dois conjuntos possuem valores diferentes para o mesmo atributo. A instabilidade se refere a variabilidade no formato, esquema e outros aspectos dos dados o que pode levar a resultados inválidos no processo de integração. A atualização frequente dos dados se torna um problema devido a frequência de atualização influenciar no rastreamento de eventos. Por exemplo, os dados meteorológicos e de tráfego são disponibilizados a cada hora, enquanto os limites administrativos estão disponíveis em um ciclo anual [Shivaprabhu et al. 2017].

Chakraborty *et al.* (2017) descrevem que o principal desafio ao usar *frameworks* ETL é a necessidade de um especialista de domínios nas fases de extração, transformação e carregamento dos dados para definir o esquema para a origem dos dados já que não existe uma ferramenta ETL capaz de mapear e definir o esquema de origem ou ontologia e carregar os dados automaticamente a partir dos metadados da fonte de dados. Os autores também relatam que *frameworks* ETL são propensos a erros devido ao ruído dos dados e afirmam que é realmente difícil definir regras na estrutura ETL para remoção de dados.

Diversos autores também relatam que tiveram problemas com banco de dados em relação as diferentes semânticas de dados [Chakraborty et al. 2017, Shivaprabhu et al. 2017, Souza et al. 2017]. Outras dificuldades enfrentadas estão relacionadas a compreensão do domínio de como geometrias GIS podem ser mapeadas e como elas podem ser exploradas em outros domínios (como gráficos), a abstração teórica de como diferentes camadas de pontos, linhas, geometrias podem ser mapeadas para multigramas, entre outros.

6. Ameaças à Validade

A fim de garantir alta qualidade e valor científico aos resultados obtidos, um protocolo foi estabelecido com as questões de pesquisa a serem respondidas e critérios explícitos para selecionar e avaliar os estudos. Mesmo assim, potenciais ameaças à validade deste estudo ainda são inevitáveis e podem afetar os resultados obtidos. Nesta seção são discutidas algumas dessas limitações e que estratégias foram adotadas para mitigá-las.

Incompletude do estudo. A principal ameaça à validade deste estudo diz respeito a sua incompletude, uma vez que estudos relevantes podem não ter sido recuperados e selecionados. Para reduzir essa ameaça, foram utilizadas três importantes bases eletrônicas de publicação dentre as existentes, porém ainda há limitações. Primeiro, alguns estudos podem não ter sido recuperados devido a limitações técnicas dos mecanismos de busca das próprias bases eletrônicas. Segundo, as bases eletrônicas escolhidas não representam uma lista exaustiva de fontes de publicação, de modo que outras bases poderiam ser

também incluídas. Terceiro, não foi realizado um processo de *snowballing*, uma técnica que consiste na verificação das listas de referências de cada um dos estudos selecionados a fim de encontrar estudos adicionais que não foram recuperados pela busca automática [Wohlin 2014]. Quarto, apesar de o fato de considerar estudos publicados entre 2015 e 2017 possibilitar focar apenas nos estudos mais recentes, essa decisão pode afetar a completude deste trabalho. Como trabalho futuro, este estudo poderá ser revisado e atualizado para cobrir essas limitações.

Incoerências e inconsistências. Para evitar possíveis incoerências e inconsistências, foi estabelecido um protocolo bem definido para orientar todo o processo de revisão da literatura. Além disso, o método de seleção dos estudos com base nos critérios de inclusão e exclusão foi escolhido com a finalidade de manter o máximo de coerência e fidelidade ao tema proposto. Esses critérios foram discutidos e planejados de maneira cuidadosa a fim de diminuir os riscos de exclusão de estudos relevantes para a pesquisa.

Viés na extração de dados. A existência de viés na extração de dados a partir dos estudos selecionados pode resultar em imprecisões, afetando assim a análise dos estudos. Mais ainda, nem todos os estudos descrevem de forma clara e suficiente as informações que seriam extraídas dos estudos para responder às questões de pesquisa, de modo que foi necessário, em alguns casos, realizar inferências durante a síntese dos resultados. A fim de minimizar tal viés, o protocolo estabelecido procurou ser seguido de maneira estrita.

7. Conclusão

A partir da revisão de literatura apresentada neste trabalho, foi possível fazer um levantamento das principais estratégias e ferramentas utilizadas no desenvolvimento de soluções de integração, relacionamento e representação de grandes volumes de dados de fontes diversas no contexto de Cidades Inteligentes. É possível observar que há um número reduzido de estudos relacionados a este tema de gerenciamento de dados e ainda há limitações nas soluções disponíveis. Foi possível observar também que as principais dificuldades enfrentadas pelos desenvolvedores são em relação a integração contínua das fontes de dados heterogêneas para suportar o desenvolvimento de aplicações e tomada de decisões.

Como trabalhos futuros, será elaborada uma estratégia baseada nos estudos encontrados nesta pesquisa na tentativa de criar uma solução de integração e gerenciamento de dados com menos limitações e capaz de integrar, relacionar e representar dados provenientes de sistemas independentes mantidos por diferentes organizações, a fim de permitir o compartilhamento de dados gerados por diferentes departamentos.

Referências

Balasubramani, B. S., Belingheri, O., Boria, E. S., Cruz, I. F., Derrible, S., Siciliano, M. D. (2017). “GUIDES – Geospatial Urban Infrastructure Data Engineering Solutions”, Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. USA: ACM.

Biolchini, J., Mian, P. G., Natali, A. C. C., Travassos, G. H. (2005) Systematic review in Software Engineering. Technical report, COPPE/Federal University of Rio de Janeiro, Brazil.

Chakraborty, J., Padki, A., Bansal, S. K. (2017) “Semantic ETL – State-of-the-art and

open research challenges”, Proceedings of the 11th IEEE International Conference on Semantic Computing. USA: IEEE, pp. 413-418

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., Rong, X. (2015) “Data Mining for the Internet of Things: Literature review and challenges”. *International Journal of Distributed Sensor Networks* 11(8).

Chen, M., Mao, S., Liu, Y. (2014) “Big Data: A survey”. *Mobile Networks and Applications* 19(2), pp. 171-209.

Kitchenham, B., Charters, S. (2007) *Guidelines for performing systematic literature reviews in Software Engineering*. Technical report, Keele University, United Kingdom.

Macêdo, J., Cacho, N., Lopes, F. (2017) “A comparative study of tools for smart cities open data publication and management”, Proceedings of the 2017 IEEE Summer School on Smart Cities. USA: IEEE – a ser publicado.

Shivaprabhu, V. R., Balasubramani, B. S., Cruz, I. F. (2017) “Ontology-based instance matching for geospatial urban data integration”, Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics. USA: ACM.

Souza, A., Pereira, J., Oliveira, J., Trindade, C., Cavalcante, E., Cacho, N., Batista, T., Lopes, F. (2017) “A data integration approach for smart cities: The case of Natal”, Proceedings of the 3rd IEEE International Smart Cities Conference. USA: IEEE.

Turner, V., Reinsel, D., Gatz, J. F., Minton, S. (2014) *The digital universe of opportunities*. IDC White Paper, EMC, USA.

Wohlin, C. (2014) “Guidelines for snowballing in systematic literature studies and a replication in Software Engineering”, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. USA: ACM.

You, L., Tuncer, B., Xing, H. (2017) “Mining place design knowledge from multi-source data in an informed design platform”, Proceedings of the 2017 IEEE International Conference on Big Knowledge. USA: IEEE, pp. 276-283.