

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

LUMA ALVES LOPES

**QUALIDADE DE DADOS E INTERNET DAS COISAS NO *FRAMEWORK* DE
RASTREABILIDADE DE GRÃOS**

PONTA GROSSA
2019

LUMA ALVES LOPES

**QUALIDADE DE DADOS E INTERNET DAS COISAS NO *FRAMEWORK* DE
RASTREABILIDADE DE GRÃOS**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para obtenção do título de Mestre.

Orientação: Prof.^a Dra. Maria Salete Marcon Gomes Vaz

PONTA GROSSA
2019

L864 Lopes, Luma Alves
 Qualidade de dados e Internet das coisas no *framework* de rastreabilidade
de grãos / Luma Alves Lopes. Ponta Grossa, 2019.
63 f.

Dissertação (Mestrado em Computação Aplicada - Área de Concentração:
Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta
Grossa.

Orientadora: Profa. Dra. Maria Salete Marcon Gomes Vaz.

1. Qualidade de dados. 2. Internet das coisas. 3. Rastreabilidade. 4.
Dimensão de qualidade. 5. Modelo de dados. I. Vaz, Maria Salete Marcon
Gomes. II. Universidade Estadual de Ponta Grossa. Computação para
Tecnologias em Agricultura. III.T.

CDD: 004

TERMO DE APROVAÇÃO

Luma Alves Lopes

QUALIDADE DE DADOS E INTERNET DAS COISAS NO FRAMEWORK DE RASTREABILIDADE DE GRÃOS

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:


Prof.^a. Dr.^a. Maria Salete Marcon Gomes Vaz
Presidente


Prof. Dr. Antônio Carlos de Francisco
UTFPR


Prof. Dr. Luciano José Senger
UEPG

Ponta Grossa, 09 de agosto de 2019.

AGRADECIMENTOS

Primeiramente, à Deus, que faz as oportunidades surgirem a todo momento, que escuta as minhas orações e que me concede a graça de superar todos os obstáculos.

À minha família, em especial aos meus pais, Robecilda e Francisco, pelo carinho, amor e esforços realizados para que eu tivesse as condições de me tornar a pessoa que sou hoje.

Agradeço à minha orientadora, Profa. Dra. Maria Salete Marcon Gomes Vaz, que aceitou me orientar nesta dissertação. Obrigada pela sua confiança, direcionamento, disponibilidade, paciência, dedicação e por permitir que eu participasse um pouco da sua rotina de pesquisa. Meu eterno respeito e admiração.

Ao meu noivo, André, pelo apoio, paciência, compreensão, carinho, amor e por toda ajuda prestada durante a elaboração desta dissertação.

Ao corpo docente do programa de Pós-Graduação em Computação Aplicada que contribuíram para o meu amadurecimento como aluna, pesquisadora e pessoa.

Um agradecimento especial aos professores Dr. Arion de Campos Junior (UEPG), Dr. Luciano José Senger (UEPG) e ao Dr. Antônio Carlos de Francisco (UTFPR) pelas contribuições na banca de qualificação e/ou defesa da dissertação.

Agradeço aos colegas do mestrado pelo apoio e companheirismo, ao Felipe e aos alunos de iniciação científica, Matheus e Rodrigo, pelos trabalhos desenvolvidos. À Leticia, que desde a graduação compartilha comigo das experiências e desafios da vida acadêmica.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior, pelo apoio financeiro que recebi durante o último ano, graças a ele pude dedicar-me exclusivamente as atividades do mestrado.

Enfim, quero demonstrar o meu agradecimento a todos aqueles que, de alguma forma ou outra, tornaram possível a realização da presente dissertação.

RESUMO

Na área agrícola há a necessidade de obter informação correta e em tempo hábil, para possibilitar o desempenho de atividades mais seguras no âmbito estratégico e operacional. Para garantir a qualidade da tomada de decisão, deve-se garantir a qualidade do dado utilizado. A utilização de dados com erros de digitação, incompletos, valores nulos, entre outros, podem resultar em relatórios que levam a tomada de decisão equivocada e/ou tardia, trazendo prejuízos para a organização. Alguns modelos de rastreabilidade permitem que seus usuários registrem, manualmente, os dados sobre os produtos e processos, o que pode gerar dúvidas sobre a qualidade do sistema e dos dados a serem rastreados. Além disso, não foi encontrada na literatura trabalhos cujo objetivo seja a qualidade de dados em sistemas de rastreabilidade de produtos agrícolas. Assim, a qualidade de dados precisa ser incorporada nas ferramentas de rastreabilidade, a fim de garantir que os seus processos são confiáveis. Esta dissertação tem por objetivo especificar um *framework* de rastreabilidade de grãos que garanta a qualidade do Atributo a ser rastreado, por meio de um modelo que considere a obtenção de dados da Internet das Coisas e critérios de qualidade de dados. Por meio do estudo de arquiteturas aplicadas na agricultura, especifica-se uma arquitetura em camadas para o *framework*. A arquitetura é avaliada conforme parâmetros de adequação à Internet das Coisas. Uma revisão da literatura sobre a percepção da qualidade de dados e a identificação de dimensões de qualidade em sistemas de rastreabilidade, definem o modelo de qualidade de dados para o *framework* de rastreabilidade de grãos. O modelo obtido é comparado com trabalhos encontrados na literatura. Como resultado, tem-se a possibilidade de coleta automática de dados, a realização de atividades de qualidade de dados para a tabela de registro sobre o Atributo rastreado, a definição de regras de qualidade para o Atributo e a possibilidade de mensurar a dimensão de qualidade de dados utilizando um método de avaliação.

Palavras-chave: Qualidade de dados. Internet das Coisas. Rastreabilidade. Dimensão de qualidade. Modelo de qualidade de dados.

ABSTRACT

In the agricultural area, accurate and timely information is needed to enable safer activities in the strategic and operational environment. To ensure the quality of decision making process, the quality of the data must be guaranteed. The use of data with typos, with incomplete information, with null values and others problems, can result in reports that lead to wrong and/or late decision making, causing damage to the organization. Some traceability models allow that users insert manual data about product and process data, which can raise questions about the quality of the system and the data to be traced. Besides, no studies were found in the literature aiming at data quality in agricultural product traceability systems. Therefore, data quality needs to be incorporated into traceability tools to ensure that their processes are reliable. This dissertation aims to specify a grain traceability framework that guarantees the quality of the Attribute to be traced, through a model that considers data obtaining from IoT and data quality criteria. Through the study of agriculturally applied architectures, we specify a layered architecture for the framework. This architecture is evaluated according to parameters of suitability to the Internet of Things. A literature review about the perception of data quality and the identification of quality dimensions in traceability systems define the data quality model for the grain traceability framework. The model obtained is compared with literature studies. As a result, we have the ability to automatically collect data, perform data quality activities for the Tracked Attribute table, define the attribute quality rules, and measure the data quality dimension, using an evaluation method.

Keywords: Data quality. Internet of Things. Traceability. Dimension of Quality. Quality Data Model.

LISTA DE FIGURAS

Figura 1 – Estrutura da dissertação.....	15
Figura 2 – Características dos dados na IoT.....	18
Figura 3 – Funcionamento do Protocolo MQTT.....	21
Figura 4 – Modelo de qualidade de dados de Cernach (2012).....	24
Figura 5 – Módulos do <i>Framework</i> RastroGrão.....	26
Figura 6 – Arquitetura de IoT proposta em camadas.....	31
Figura 7 – Fluxo de informação na camada de dispositivos para obtenção de dados.....	32
Figura 8 – Fluxo de informação do subprocesso “Realizar controle da Qualidade de dados”.....	33
Figura 9 – Troca de mensagens entre Publicadores, Broker e Assinantes.....	34
Figura 10 – Modelo de domínio.....	37
Figura 11 – Estrutura de Customização para o <i>Framework</i>	38
Figura 12 – Modelo de informação.....	39
Figura 13 – Modelo funcional.....	40
Figura 14 – Extensão da estrutura de rastreabilidade de grãos para a integração da IoT.....	45
Figura 15 – Modelo de Qualidade de dados para a entidade “Lot”.....	49
Figura 16 – Modelo de Qualidade de dados para o <i>Framework</i> de rastreabilidade de grãos.....	51
Figura 17 – Extensão da estrutura de rastreabilidade de grãos para a gestão da QD.....	55

LISTA DE QUADROS

Quadro 1 – Etapas de desenvolvimento da pesquisa.....	28
Quadro 2 – Avaliação das arquiteturas.....	43
Quadro 3 – Dimensões de qualidade de dados identificados em sistemas de rastreabilidade..	53

LISTA DE ABREVIATURAS E SIGLAS

BPMN	Notação de Modelo de Processos de Negócio
COAP	<i>Constrained Application Protocol</i>
DTN	<i>Delay Tolerant Networks</i>
ER	Entidade-Relacionamento
FT-IR	Espectroscopia de infravermelho por transformada de Fourier
GS-SVM	Grid Search Support Vector Machine
IOE	<i>Internet of Everything</i>
IOT	Internet das Coisas
IOT-A	<i>Internet of Things Architecture</i>
IP	<i>Internet Protocol</i>
M2M	Máquina para Máquina
MQTT	<i>Message Queue Telemetry Transport</i>
Q	<i>Quality</i>
QD	Qualidade de dados
QFR	<i>Quality Fault Resolve</i>
QR-CODE	<i>Quick Response Code</i>
QTO	<i>Quality Test Operation</i>
RFID	Identificação por Rádio Frequência
RSSF	Redes de Sensores sem Fio
UML	Linguagem de Modelagem Unificada
UV-VIS	Espectroscopia no ultravioleta visível

SUMÁRIO

1	INTRODUÇÃO	11
1.1	PROBLEMÁTICA E MOTIVAÇÃO	11
1.2	JUSTIFICATIVA	12
1.3	OBJETIVOS	14
1.4	ESTRUTURA DO TRABALHO	14
2	REVISÃO DA LITERATURA	16
2.1	CONCEITOS	16
2.2	INTERNET DAS COISAS	17
2.2.1	Arquitetura na Internet das Coisas	20
2.2.2	Protocolos para Troca de Mensagens	21
2.3	QUALIDADE DE DADOS	22
2.4	RASTREABILIDADE	25
3	METODOLOGIA.....	28
3.1	MATERIAL E MÉTODOS	28
3.2	CARACTERIZAÇÃO DA PESQUISA	30
4	ESPECIFICAÇÃO DE UMA ARQUITETURA DE IOT PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS	31
4.1	ARQUITETURA BASEADA EM CAMADAS	31
4.1.1	Camada de Dispositivos	32
4.1.2	Camada de Comunicação	33
4.1.3	Camada de Serviço	35
4.1.4	Camada de Dados	35
4.1.5	Camada de Aplicação	35
4.1.6	Outros Componentes da Arquitetura	35
4.1.6.1	Segurança	35
4.1.6.2	Qualidade de dados.....	36
4.2	MAPEAMENTO DA SOLUÇÃO NOS MODELOS PROPOSTOS NA IOT-A.....	36
4.2.1	Modelo de Domínio.....	36
4.2.2	Modelo de Informação.....	38
4.2.3	Modelo Funcional.....	39
4.3	TRABALHOS CORRELATOS	40
4.4	AVALIAÇÃO DA ARQUITETURA	42
4.5	CONSIDERAÇÕES FINAIS	44
5	UM MODELO DE QUALIDADE DE DADOS PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS	46
5.1	QUALIDADE DE DADOS EM SISTEMAS DE RASTREABILIDADE DE PRODUTOS AGRÍCOLAS	46
5.1.1	Coleta Automática de Dados	46
5.1.2	Fusão de Dados.....	47
5.1.3	Documentação e Modelagem dos Dados	47
5.1.4	Especificações de Ontologias e Semântica.....	48

5.1.5	Modelos de Qualidade de Dados	48
5.2	DEFINIÇÃO DA ESTRUTURA DE QUALIDADE DE DADOS PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS	49
5.2.1	Fase de Planejamento	50
5.2.2	Fase de Preparação	50
5.3	IDENTIFICAÇÃO DE DIMENSÕES DE QUALIDADE DE DADOS	52
5.4	CONSIDERAÇÕES FINAIS	54
6	CONCLUSÕES E PESQUISAS FUTURAS.....	57
	REFERÊNCIAS	59

1 INTRODUÇÃO

Este Capítulo, na seção 1.1, traz a problemática e a motivação do tema deste trabalho. Em seguida, na seção 1.2, são apresentadas as justificativas encontradas para a realização desta dissertação. Na seção 1.3 são discutidos os objetivos, geral e específicos. Por fim, a seção 1.4 apresenta a organização desta dissertação.

1.1 PROBLEMÁTICA E MOTIVAÇÃO

A necessidade de proporcionar segurança alimentar e evitar problemas como intoxicação humana por agentes patogênicos, motiva a criação de sistemas e processos de rastreabilidade.

Alguns modelos de rastreabilidade de grãos, como o de Ceruti (2007) e o de Vaz (2014), permitem que os usuários customizem as suas estruturas de inserção de dados conforme as suas necessidades, ou seja, o registro dos dados sobre os produtos e processos dentro da cadeia produtiva, fica a cargo de seus usuários. Entretanto, estes modelos não abordam a coleta automática de dados e, uma vez que os dados sejam inseridos manualmente, podem gerar dúvidas sobre a qualidade do sistema e dos dados a serem rastreados.

No *framework* RastroGrão, especificado por Vaz (2014) com o objetivo de rastrear qualquer tipo de grão, por meio do gerenciamento de Produtos, Fases de produção e dos Atributos a serem rastreados, os registros de dados sobre os Atributos representam o principal suporte para a tomada de decisão. A baixa qualidade desses dados, pode influenciar negativamente na informação e no conhecimento extraído da base de dados e, por tanto, comprometer a qualidade dos atributos rastreados e do sistema de rastreabilidade.

Como os dados representam o principal suporte para a tomada de decisão, a existência de dados ruins, sejam por erros de digitação, dados incompletos, valores nulos, entre outros, podem resultar em relatórios que levam a tomadas de decisão erradas e/ou tardias, trazendo prejuízos para a organização.

Na área agrícola é importante a obtenção de uma informação correta e em tempo hábil, para possibilitar que sejam desempenhadas atividades mais seguras no âmbito estratégico e operacional. Portanto, um processo de rastreabilidade eficiente deve garantir que os dados a serem rastreados estejam em conformidade com a realidade.

Nesse contexto, com a Internet das Coisas (IoT) e a disponibilidade dos dados em tempo real, pode-se ter uma tomada de decisão precisa e confiável ao possibilitar que

ferramentas inteligentes coletem os dados automaticamente e enviem para o sistema de rastreabilidade. A utilização no monitoramento e rastreamento dos produtos em seus ciclos de vida, beneficia os consumidores, ao possibilitar a obtenção de informações detalhadas sobre a origem e/ou métodos de produção dos produtos.

O potencial da IoT está em disponibilizar, em tempo real, dados de diferentes fontes (AGGARWAL et al., 2013). Um dos desafios está no tratamento dos dados que, mesmo sendo coletados automaticamente, ainda podem conter problemas que, se não tratados, podem influenciar negativamente a tomada de decisão. Esse problema pode ser minimizado, ao utilizar mecanismos para detecção e correção de problemas, visando a qualidade de dados (OLIVEIRA, 2008).

Dessa forma, a introdução da IoT e de um modelo de qualidade de dados em um *framework* de rastreabilidade de grãos, permite que os dados armazenados no banco de dados e disponíveis para a visualização ao usuário, tenham sido passados por critérios, a fim de garantir a qualidade da informação extraída do sistema.

1.2 JUSTIFICATIVA

Com o aumento da população e a demanda por alimentos de qualidade, surge a necessidade de modernização e intensificação das práticas agrícolas. Nesse sentido, a IoT na agricultura contribui no controle da qualidade da produção de alimentos de forma sustentável.

As vantagens na integração da IoT na agricultura, envolvem a otimização da produção, minimização de custos, aumento da qualidade, da confiabilidade e da valorização dos produtos, dentre outras. Revisões sistemáticas da literatura científica a partir do ano de 2010, mostram o crescente interesse na utilização da IoT na agricultura. (KHANNA; KAUR, 2019; TZOUNIS et al., 2017).

Dentre os exemplos de aplicação da IoT na agricultura, tem-se a rastreabilidade da cadeia produtiva. Os dados podem ser coletados por diferentes objetos que também podem monitorar toda a cadeia produtiva. A inserção dos dados possibilita comprovar a origem do produto, proporciona segurança alimentar, além de proporcionar qualidade ao produto.

Assim, torna-se fundamental garantir a qualidade dos dados nesses sistemas. Essa qualidade dos dados pode ser introduzida por meio de entidades de qualidade que, ao serem introduzidas no banco de dados juntamente com as demais entidades da aplicação, compõem um modelo conceitual de qualidade de dados (QD) que pode ser convertido em um modelo de Entidade-Relacionamento (ER) (BATINI; SCANNAPIECO, 2006).

Sob o ponto de vista de negócios, uma pesquisa realizada com 50 empresas de diferentes setores, mostrou que 58,8% dos entrevistados alcançaram resultados benéficos e mensuráveis após investirem em Big Data, sendo que 80,7% dos entrevistados classificaram o investimento como sendo bem-sucedido. O retorno do investimento inclui a redução de custos, o estabelecimento de uma cultura baseada em dados, além de benefícios relacionados à inovação de produtos e serviços (NEW VENTAGE PARTNERS, 2017).

Mesmo com os investimentos, as empresas se deparam com base de dados inconsistentes que prejudicam suas estratégias de negócio. Em uma pesquisa realizada pela empresa Piattino (2017) sobre a qualidade dos dados nos resultados das ações das empresas, mostra que a maioria dos entrevistados atribui problemas de qualidade de dados a erros humano e falta de conhecimento. Os principais erros são referentes a dados incompletos (60%), informações desatualizadas (54%) e dados duplicados (51%).

Por meio dos dados é possível obter informações que poderão ser utilizadas para realizar tomadas de decisões e elaborar planos estratégicos para as organizações. Manter dados com problemas, afeta a qualidade das informações, do conhecimento e das tomadas de decisões. Dessa forma, destaca-se a importância de dados de qualidade incorporados nos sistemas utilizados pelas organizações.

Considerando a importância dos dados para as organizações, a qualidade dos dados deve ser obrigatória em sistemas de rastreabilidade de grãos. Entretanto, a realização de uma revisão da literatura entre os anos de 2010 e 2018, utilizando as palavras-chave *avaliação e sistema de rastreabilidade, avaliação e framework de rastreabilidade, qualidade e dado e rastreabilidade, qualidade e informação e rastreabilidade, qualidade e dado e agricultura, qualidade e dado e alimento, qualidade e dado e grãos, qualidade e dado e frutas*, mostrou que não existem pesquisas cujo o objetivo geral seja a qualidade dos dados em modelos de rastreabilidade de produtos agrícolas, embora seja possível observar percepções de qualidade de dados.

Dessa forma, considera-se importante especificar um modelo de rastreabilidade de grãos que se preocupe com a qualidade dos dados sobre os Atributos a serem rastreados, pois contribui para a melhoria dos modelos e processos organizacionais, além de proporcionar ao consumidor dos dados tomadas de decisão mais seguras e assertivas, uma vez que as informações e conhecimento extraído por meio desses dados, se tornam mais confiáveis.

1.3 OBJETIVOS

O objetivo geral dessa dissertação é especificar um *framework* de rastreabilidade de grãos que garanta a qualidade do Atributo a ser rastreado, por meio de um modelo de qualidade de dados que considere a obtenção de dados da Internet das Coisas e critérios de qualidade de dados.

Os objetivos específicos são os que seguem:

- Possibilitar que os dados sejam coletados automaticamente; e
- Contemplar dimensões de qualidade de dados para a tabela de registro de dados.

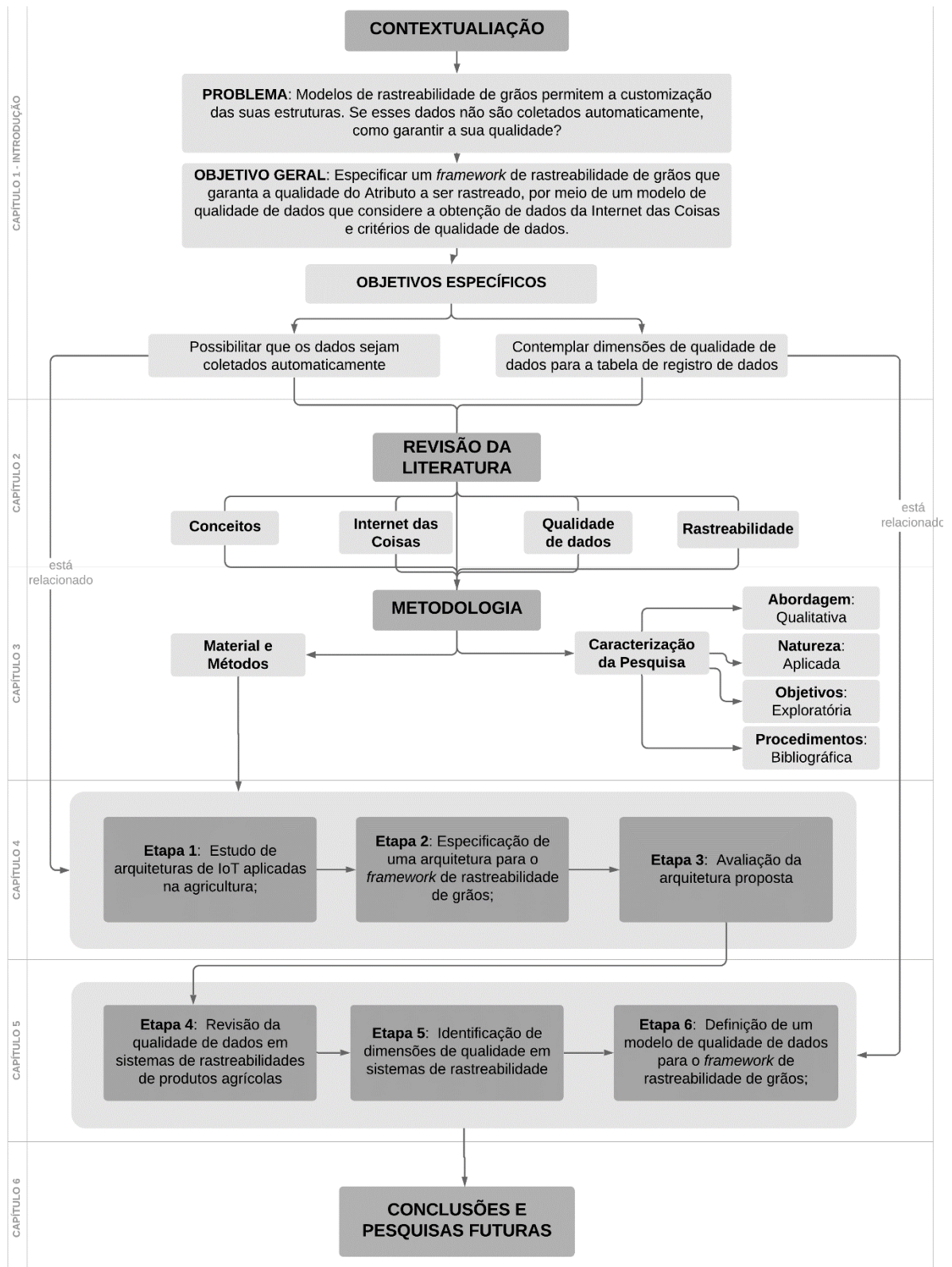
1.4 ESTRUTURA DO TRABALHO

Esta dissertação está organizada em seis capítulos. O Capítulo 2 aborda a fundamentação teórica sobre a Internet das Coisas, Qualidade de dados e Rastreabilidade. O Capítulo 3, material e método, descreve como o trabalho foi executado, o material utilizado e os métodos empregados.

O Capítulo 4 descreve a arquitetura de IoT especificada, enquanto que o Capítulo 5 descreve o modelo de qualidade de dados para o *framework* de rastreabilidade de grãos. Por fim, o Capítulo 6, descreve as considerações finais.

Os passos realizados para o desenvolvimento desta dissertação, está delineado conforme a Figura 1.

Figura 1 – Estrutura da dissertação



Fonte: A Autora

2 REVISÃO DA LITERATURA

Nesta seção é abordada a fundamentação teórica para essa proposta de dissertação, envolvendo Internet das Coisas, qualidade de dados e rastreabilidade. A Subseção 4.1 descreve os conceitos que envolve a Internet das Coisas. A Subseção 4.2 descreve as dimensões, preocupações e como avaliar um projeto de qualidade de dados. Finalmente, a Subseção 4.3 descreve os conceitos de rastreabilidade e o *framework* de rastreabilidade de grãos.

2.1 CONCEITOS

Para a entendimento da arquitetura proposta, tendo como referência Walewski et al. (2013) e Oliveira (2017), torna-se necessário a discussão dos conceitos:

- *Dispositivo*: Considera como sendo um dispositivo, qualquer *hardware* utilizado para capturar dados ou interagir com o meio ambiente. Um Dispositivo pode disponibilizar mais de um Recurso. Exemplo de Dispositivos: Sensores e atuadores ou outros dispositivos como, por exemplo, narizes eletrônicos que podem ser implementados em armazéns de grãos para verificar a qualidade do produto armazenado.
- *Recurso*: Recursos são as funcionalidades que possibilitam obter dados ou ativar um atuador. Os Recursos podem ser disponibilizados pelos dispositivos ou pela rede. Em caso de utilização de recursos da rede, esses podem se referir aos casos onde realiza-se a conexão com outros *softwares*, ou se utilize os recursos do próprio *framework*. Cada Recurso é identificado unicamente para garantir a integridade do sistema. Exemplo: Iniciar Leitura, obter temperatura ou ativar a irrigação.
- *Serviço*: Entende-se por serviço uma funcionalidade de negócio. No contexto de *hardware*, um serviço é disponibilizado por um Dispositivo por meio dos seus Recursos. Exemplo: Monitorar uma Entidade Física.
- *Entidade Física*: São representações do mundo real que podem ser gerenciadas por Dispositivos. Exemplo: Um Lote, pode ser considerada uma entidade física, em que sua qualidade pode ser constatada por meio da utilização de sensores.
- *Tópico*: Identifica o assunto das informações que serão publicadas. No sistema de rastreabilidade, refere-se aos Atributos que serão rastreados durante cada fase do processo produtivo do produto.

- *Publicações*: As publicações são as mensagens enviadas por um Dispositivo para serem processadas em um tópico específico.

2.2 INTERNET DAS COISAS

Na Internet das Coisas, os computadores são capazes de obter informações sem interferência humana, ao utilizar tecnologias que os permitem observar, identificar e compreender o mundo (ASHTON, 2009).

No ambiente IoT há presença de objetos como Identificação por Rádio Frequência (RFID), sensores, atuadores, celulares e etc. que, por meio de mecanismos de endereçamento, como a Internet, interagem e cooperam entre si para atingir algum objetivo (GIUSTO et al., 2010).

A definição de IoT, também conhecida por Internet de Tudo (*Internet of Everything - IoE*) (HUSSAIN, 2017), direciona para uma rede de comunicação, envolvendo processos, sociedade, fenômenos, dados e objetos. Esse conceito possibilita, do ponto de vista de dados e objetos, que dados sejam transformados em informações corretas e, uma vez que sejam entregues aos dispositivos e objetos interligados no instante adequado, permite tomadas de decisões inteligentes.

Com base nesses conceitos de IoT, pode-se ter sensores interconectados criando uma Rede de Sensores sem fio - RSSF (ALABA et al., 2017). Uma RSSF corresponde ao uso de sensores geograficamente distribuídos em um ambiente, capazes de monitorar o meio, coletar e processar os dados e, em seguida, enviá-los a um ou mais pontos de interesse por meio da comunicação sem fio (DWIVEDI; VYAS, 2011).

A utilização de RSSF na IoT, visa substituir o monitoramento via um único sensor, já que por ser único, necessita de uma alta confiabilidade na mediação, baixa tolerância a falhas, alto custo de manutenção, além de não representar o ambiente como um todo. Um dos principais desafios da RSSF está relacionado a análise dos dados que são transmitidos.

Nesse sentido, a IoT é importante para a economia e para a sociedade (LI et al., 2012) e tem o potencial para desenvolvimento de aplicações inteligentes em diversos domínios, como na indústria, em cidades inteligentes, e na área de saúde e bem-estar (BORGIA, 2014).

No domínio da indústria pode-se subdividi-la em três áreas: agricultura e pecuária, logística e gerenciamento de produtos e processo industrial. Já em cidades inteligentes temos mobilidade e turismo, redes inteligentes, edifícios e casas inteligentes. Por fim, em saúde e bem-estar pode-se subdividi-la em duas áreas: saúde pública e meio ambiente, medicina e saúde.

Na agricultura, alguns exemplos de aplicação da IoT são no monitoramento do clima e do solo, otimização de irrigação, controle de estufas, controle do efeito estufa e na rastreabilidade da cadeia produtiva (TZOUNIS et al., 2017).

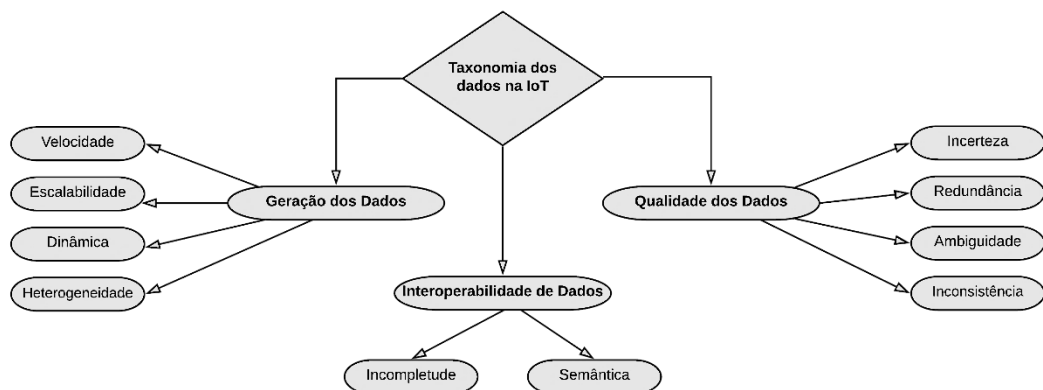
Na rastreabilidade, os objetos da IoT podem monitorar toda a cadeia produtiva, coletando dados de sensores, drones, imagens de satélites, máquinas automatizadas, entre outras, desde o plantio e/ou semeadura até a entrega do produto ao consumidor final.

Dessa forma, tem beneficiado os consumidores, ao permitir, por exemplo, que acessem informações adicionais sobre os produtos, além dos fabricantes, ao aumentar o número de produtos que utilizam tecnologias de identificação como Etiquetas RFID e sensores inteligentes, possibilitando que os produtos sejam monitorados e rastreados em seus ciclos de vida.

O potencial da IoT está em disponibilizar, em tempo real, dados de diferentes fontes para uma variedade de serviços. Um dos desafios é como tratar esses dados que podem conter muitos ruídos, em grande escala e de forma distribuída (AGGARWAL et al., 2013).

Conforme ilustra a Figura 2, as características dos dados na IoT podem ser classificadas nas categorias: Geração de Dados, Interoperabilidade de Dados e Qualidade de Dados (QIN et al., 2016).

Figura 2 – Características dos dados na IoT



Fonte: Traduzido de QIN et al. (2016)

A geração de dados da IoT preocupa-se com os fatores velocidade, escalabilidade, dinâmica e heterogeneidade. Com relação a velocidade, altas taxas de amostragem exigem um processamento eficiente, mas baixas taxas podem provocar perdas de informações importantes para o processamento e tomada de decisão. A alta quantidade de dados que é gerada pelas coisas

conectadas, exige que os sistemas de processamento dos dados sejam escaláveis e possuam mecanismos para tratamento desses dados.

O dinamismo dos dados é importante para a geração de dados já que, como muitas coisas são móveis, acabam estando em lugares diferentes em momentos diferentes. Além disso, ao longo do tempo, os dados podem sofrer por falhas nas coisas. A heterogeneidade de coisas diferentes gerando dados de diferentes formatos, resulta na preocupação em atender as diferentes necessidades de dados.

A interoperabilidade está relacionada a incompletude e semântica dos dados. No cenário da IoT, é necessário combinar diferentes tipos de fonte, a fim de determinar quais delas podem atender melhor a incompletude de dados para uma tarefa de processamento. Já a semântica dos dados refere-se em permitir às coisas/máquinas entender e processar os dados.

Finalmente, a qualidade dos dados está relacionada a incerteza, redundância, ambiguidade e inconsistência. Essas características estão relacionadas, por exemplo, ao tratamento de leituras ausentes, de leituras semelhantes para uma mesma área, de interpretação dos dados de acordo com as necessidades de cada consumidor desses dados, de problemas na detecção de dados ou de perda de pacotes durante a transmissão. Esses dados podem ser gerados automaticamente ou por meio de uma ação humana, e podem ser: (COOPER; JAMES, 2009):

- RFID refere-se a identificação e rastreamento, usando ondas de rádio. Etiquetas RFID são utilizadas para transmitir e receber informações sobre os objetos;
- Endereços ou identificadores únicos dos objetos;
- Dados descritivos dos objetos, processos e sistemas;
- Dados posicionais ou da localização geográfica;
- Dados de sensores;
- Dados históricos;
- Modelos físicos como, por exemplo, gravidade, força, luz, som e magnetismo; e
- Estado dos Atuadores e Dados de comandos para controle.

Para especificar uma estrutura de rastreabilidade de grãos baseada em IoT, é necessário definir uma arquitetura e um protocolo de comunicação entre os objetos nesse cenário. Por isso, as subseções 2.1.1 e 2.1.2 descrevem, respectivamente, a arquitetura na IoT e os protocolos para a troca de mensagens/dados entre os objetos na internet das coisas.

2.2.1 Arquitetura na Internet das Coisas

Um modelo de referência é um artefato que descreve conceitos e relacionamentos em um domínio específico. A arquitetura de referência serve como base para a construção e projeto de arquiteturas com base em modelos de referência (NAKAGAWA et al., 2014).

Na literatura são encontradas diversas arquiteturas para a compreensão dos conceitos de IoT (SETHI; SARANGI, 2017). Dentre essas arquiteturas, temos Projeto *Internet of Things Architecture* (IoT-A), que provê um modelo e uma arquitetura de referência para a IoT (WALEWSKI et al., 2013).

O modelo de referência IoT-A, foi proposto para descrever a IoT incluindo: um modelo de domínio, que descreve os conceitos pertencentes a uma área de interesse; o modelo de informação, responsável pela descrição dos dados a serem processados; e o modelo funcional, que descreve a interação entre os objetos (WALEWSKI et al., 2013).

O Modelo de Domínio é independente de tecnologia e introduz os conceitos de IoT, como dispositivos, serviços, recursos, entidades virtuais, os atributos básicos e as relações entre os objetos.

O Modelo de Informação preocupa-se com a estrutura que contém os dados a serem manipulados em um sistema de IoT, sem a preocupação em como deve ser representado.

O Modelo Funcional identifica grupos funcionais como, por exemplo, o Modelo de Comunicação e o Modelo de Segurança. Esses grupos provem as funcionalidades para interação com as instâncias dos conceitos do Modelo de Domínio, ou gerenciam as informações relacionadas aos conceitos, usando como base o Modelo de Informação.

O Modelo de Comunicação é um grupo funcional que introduz conceitos para a manipulação da comunicação necessária entre os componentes no ambiente da IoT. Além disso, como a segurança é importante na IoT, as funcionalidades e interações relevantes são introduzidas no Modelo de Segurança.

Além de um modelo de referência, o projeto IoT-A provê uma arquitetura de referência, com base em visões arquiteturais para a construção de diferentes arquiteturas para IoT. As visões propostas são: Visão Funcional, Visão de Informação, e Visão de desenvolvimento e Operação (WALEWSKI et al., 2013).

A Visão funcional é caracterizada pelo mapeamento dos requisitos; funcionalidades semelhantes são agrupadas e componentes funcionais são definidos, sem se preocupar com as interações entre os elementos, já que dependem das escolhas de projeto.

O principal propósito dos objetos conectados na IoT, é a troca de informações. Por isso, a Visão Informacional, enfoca-se na descrição, tratamento, ciclo de vida e fluxo das informações. Já a Visão de desenvolvimento e operação, tem por objetivo fornecer um conjunto de orientações para guiar usuários na escolha de projetos. Descreve serviços, elementos funcionais, e tecnologias disponíveis na IoT para a construção e desenvolvimento das aplicações.

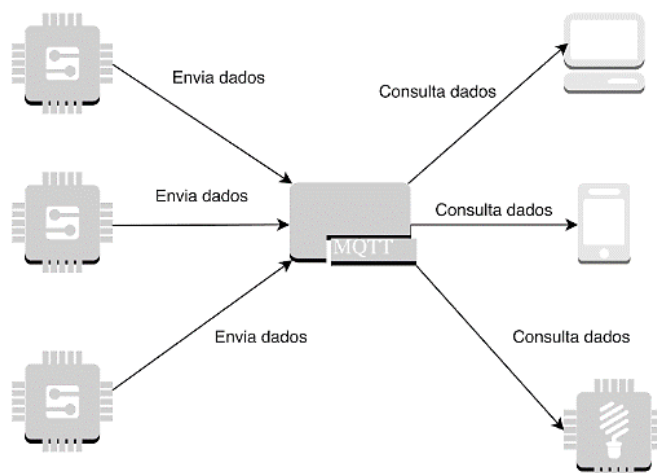
2.2.2 Protocolos para Troca de Mensagens

A camada de aplicação da IoT usa os dados processados para implementar as suas aplicações. Dessa forma, existe a necessidade de pensar em um protocolo que gerencie a troca de mensagens e/ou dados entre os elementos que compõe a rede. Nesse cenário tem-se, por exemplo: Message Queue Telemetry Transport (MQTT) e Constrained Application Protocol (CoAP).

O MQTT é um protocolo baseado na arquitetura *publish/subscribe* e *broker* para transferência de mensagens. Dentre os seus princípios, tem-se a minimização de utilização dos recursos dos dispositivos e da largura de banda, tentando garantir confiabilidade e garantia de entrega da mensagem (MQTT, 2019).

O funcionamento do protocolo tem três elementos: o *Broker*, servidor que faz a intermediação entre a geração de dados e sua publicação; o *Publisher*, dispositivos que geram e enviam dados para o *broker*; e *Subscriber*, o assinante do serviço disponibilizado pelo *publisher*. A Figura 3 ilustra o funcionamento do protocolo MQTT (OLIVEIRA, 2017).

Figura 3 – Funcionamento do Protocolo MQTT



No MQTT, os dispositivos realizam a inscrição (*subscribe*) em um *broker*, a fim de realizarem consultas de informações específicas (tópicos, info), sempre que publicadores (*publishers*) divulgarem os dados de interesse, por meio do *broker*.

Outro exemplo de protocolo é o CoAP, que foi desenvolvido para dispositivos com baixa capacidade de processamento e armazenamento. O seu modelo de funcionamento é baseado em requisição/resposta, dando suporte a descoberta de serviços e foi projetado para interagir com o Protocolo de Transferência de Hipertexto (HTTP) na Web. Algumas das suas características são, por exemplo, suporte à *Uniform Resource Identifier* (URIs), e suporte aos métodos GET, POST, PUT e DELETE (SHELBY et al. 2014).

2.3 QUALIDADE DE DADOS

A definição de qualidade de dados pode variar de acordo com a sua intenção de uso. Geralmente, os dados devem satisfazer determinados critérios para serem considerados de alta qualidade. Dessa forma, a qualidade de dados pode ser definida como um valor agregado a um conjunto de critérios de qualidade (AHMED; AZIZ, 2010).

Dentre os critérios de qualidade, pode-se destacar as seguintes dimensões da qualidade dos dados (OLSON, 2003):

- *Acurácia*: Os valores dos dados armazenados devem estar corretos. Existem duas características fundamentais, formato e o conteúdo;
- *Atualidade*: Os dados devem ser disponibilizados no tempo desejável;
- *Relevância*: Devem ser importantes para o domínio de aplicação;
- *Completude*: Todos os dados devem ser disponibilizados de acordo com um contexto de utilização;
- *Compreensibilidade*: Permitir o uso correto dos dados por meio do entendimento do seu significado;
- *Confiabilidade*: Tanto os dados quanto a aplicação que os armazena e manipula, devem ser confiáveis.

No controle de dados sobre um lote, por exemplo, é importante a existência de acurácia nos dados armazenados no sistema. Para que exista acuracidade no controle lógico do lote é necessário que todos os dados gerados, durante a execução dos processos, tenham sido registrados de forma correta no sistema de rastreabilidade, além de estarem exatamente

compatíveis com o processo real realizado. Se, por exemplo, o lote é registrado no sistema com a data de fabricação de 16/05/2019 e o correto seria 14/05/2019, existe acurácia nesse dado quanto ao formato, mas não existe acurácia quanto ao seu conteúdo.

Ainda considerando o exemplo do controle de lotes, se existir a necessidade da realização do monitoramento durante o transporte, um dado sobre a localização geográfica em tempo real deve ser mantido. Dessa forma, pode-se dizer que esse dado possui completude e relevância, por estar sendo utilizado de acordo com um contexto e ser considerado importante para o domínio, mas também possui compreensibilidade, já que se entende o contexto de utilização desse dado. Se o consumidor do dado desejar ter acesso a essa localização a cada 60 segundos, para que esse dado possua atualidade, esse intervalo de tempo deve ser respeitado.

Com os dados sendo inseridos, armazenados e gerenciados com acurácia, torna possível a existência da confiabilidade dos dados e do sistema. Essa confiabilidade está relacionada com a percepção e avaliação da potencialidade de uso desse dado de acordo com a finalidade de um determinado contexto.

A preocupação com a qualidade dos dados, envolvem: Detecção e Correção de problemas como, por exemplo, valores em falta ou registros duplicados; Migração de dados pouco estruturados para uma base de dados estruturada, para facilitar a manipulação; Integração de dados provenientes de diversas fontes em uma nova fonte de dados (GALHARDAS et al., 2000).

Existem diversos fatores que influenciam a qualidade dos dados. Dentre os fatores que são consequência da ação humana, podemos citar: erros na digitação de valores, a omissão de valores em atributos supostamente obrigatórios, a introdução involuntária de valores errados causados por formulários confusos, entre outros. Dentre os fatores que influenciam a qualidade dos dados, que não dependem diretamente de um erro humano, envolvem: erros de medição por um sensor defeituoso, erro na transmissão dos dados, e erro na transformação ou integração dos dados (OLIVEIRA, 2008).

Para realizar a melhoria da qualidade dos dados, torna-se necessária a realização de uma ou mais atividades de detecção e de correção de problemas, sendo que pelo menos uma das atividades seja de correção. As atividades de detecção que, eventualmente, podem ser realizadas são: *Data Profiling* e análise de dados. Já as de correção são: transformação de dados, limpeza de dados e enriquecimento de dados (OLIVEIRA, 2008).

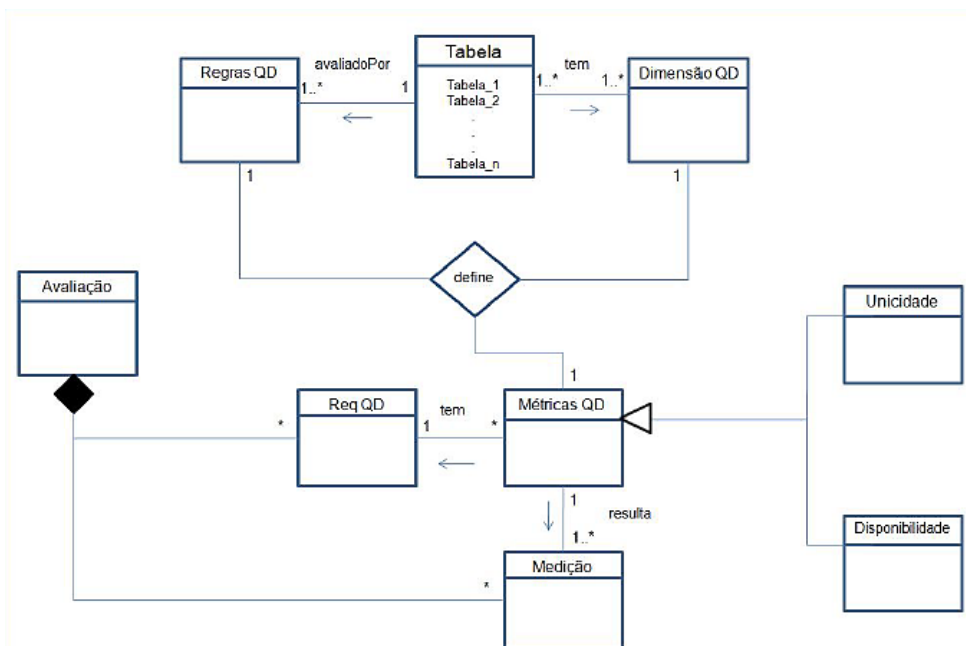
Na atividade de *Data Profiling* aplica-se técnicas de análise para determinar o conteúdo, estrutura e qualidade dos dados existentes, sendo que o conhecimento gerado é sobre

os próprios dados. Já na análise de dados infere-se padrões, regras e relacionamentos e envolve as tarefas de inferência de estrutura e de detecção de anomalias.

A atividade de transformação de dados, envolve operações de modificações no esquema dos dados, a fim de que um esquema de origem fique de acordo com um esquema de destino. Na limpeza de dados, são utilizadas técnicas que envolve remoção de registros com problemas, atribuição de valores padrões, ou aplicação de técnicas de agrupamento para auxiliar na descoberta dos novos valores. Já a atividade de enriquecimento de dados, consiste em melhorar a precisão e completude dos dados, com base em dados existentes em outras fontes.

A qualidade dos dados pode ser incorporada por meio de modelos de qualidade de dados. Em Cernach (2012) no modelo especificado (Figura 4), aplicado na provisão de custos com base nos riscos calculados provenientes de contratos de crédito não quitados, a avaliação de tabela do banco de dados considera: regras de qualidade de dados, dimensões de qualidade e métricas de avaliação, em que se avaliou as métricas Unicidade e Disponibilidade.

Figura 4 – Modelo de qualidade de dados de Cernach (2012)



Fonte: CERNACH (2012)

Para avaliar um projeto de qualidade de dados, pode-se realizar quatro fases (MAYDANCHIK, 2007):

- *Fase de Planejamento*: define-se o escopo a ser estudado, conhece-se as tabelas, registros, campos que são relevantes e que podem prover definições de qualidade de dados;

- *Fase de Preparação*: compreende o desenvolvimento de catálogos e modelos de dados, ou seja, uma coleção de ferramentas para organização de todos os dados;
- *Fase de Implementação*: antes de especificar as regras de qualidade, que são a ferramenta principal de um projeto de qualidade de dados, realiza-se uma atividade de *data profiling*. As regras que validam os dados e os relacionamentos são desenvolvidos por meio de linguagens de programação. Essas regras envolvem restrição de valores de atributos, regras de integridade relacional, regras para dados históricos e regras de dependência geral;
- *Fase de Refinamento das Regras de Qualidade*: as regras são validadas por especialistas, a fim de aprimorá-las e atingir o máximo da precisão com relação à identificação de erros.

No trabalho de Cernach (2012), a avaliação do modelo de qualidade de dados contempla as fases de planejamento, preparação e implementação, conforme encontrado em Maydanchik (2007).

2.4 RASTREABILIDADE

A rastreabilidade pode ser compreendida como a habilidade de acompanhar, rastrear e identificar unicamente um produto em todas as suas etapas de produção, processamento e distribuição (DERRICK; DILON, 2004).

Por meio do acesso aos registros, das informações e manuseio dos dados em qualquer ponto da cadeia produtiva, a rastreabilidade garante a qualidade dos produtos e proteção aos consumidores, ao possibilitar a eliminação de produtos alimentícios defeituosos e promover a investigação das causas dos problemas (BADIA-MELIS et al., 2015).

Um sistema de rastreabilidade orientado pela qualidade, além de dados sobre localização e identificação, preocupa-se com dados relevantes sobre a qualidade do produto como, por exemplo, umidade e temperatura, que são capturados por meio de ferramentas inteligentes, como sensores, para controlar e gerenciar o fluxo dos produtos na cadeia produtiva (SCHEER, 2006).

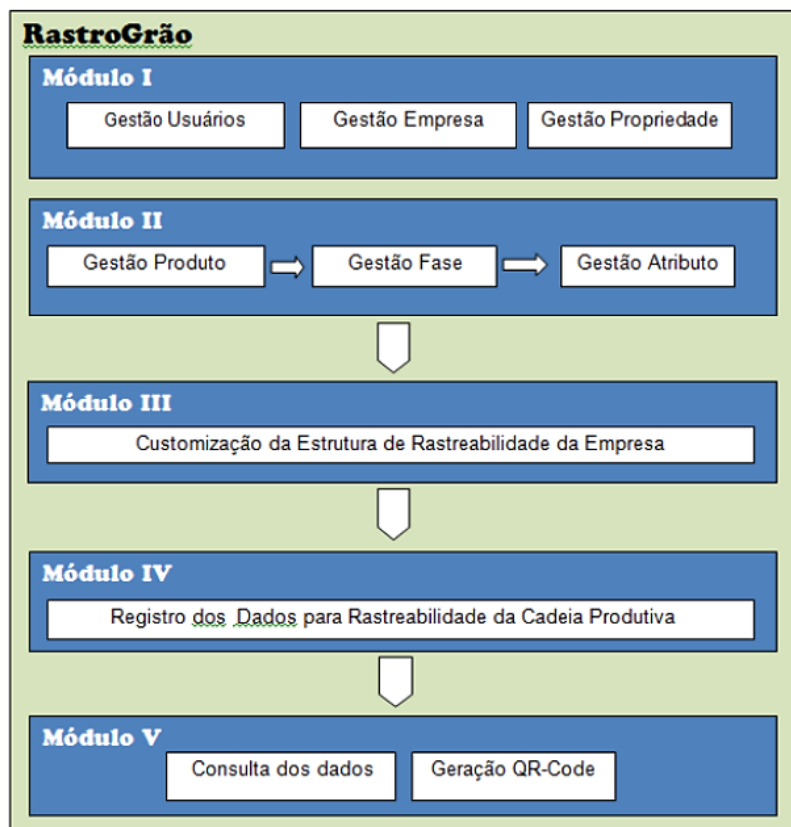
Dessa forma, pode-se destacar que um sistema de rastreabilidade eficiente, também exposto por Gazolla et al. (2017), tem por finalidade garantir segurança alimentar, controle de origem e de qualidade do produto, direito do consumidor à informação, e utiliza inovações tecnológicas para aperfeiçoar a organização das cadeias produtivas.

A ISO (2007) destaca que um sistema de rastreabilidade deve considerar a existência de outros sistemas de operação e gerenciamento presentes em uma organização.

O *framework* de rastreabilidade, RastroGrão, foi especificado por Vaz (2014) com o objetivo de rastrear qualquer tipo grão, por meio do gerenciamento de Produto, Fases de Produção e dos Atributos a serem rastreados em cada fase de produção.

O acesso ao sistema é realizado pela internet, permitindo que as informações sejam acessadas e customizadas por todos os agentes da cadeia já que, a estrutura do *framework* dividido em cinco módulos (Figura 5), permite a adaptação às regras de cada agronegócio.

Figura 5 – Módulos do *Framework* RastroGrão



Fonte: VAZ (2014)

O Módulo I, caracteriza-se pela gestão dos Usuários que terão acesso ao sistema, das Empresas envolvidas no processo de rastreabilidade e das Propriedades onde o grão é produzido. Já o Módulo II, compreende a gestão da estrutura da rastreabilidade, criando os Produtos, Fases dos produtos e os Atributos a serem rastreados.

O Módulo III é dependente do Módulo II, por ser responsável pela customização da estrutura de Rastreabilidade da Empresa, onde são selecionados os produtos, fases e atributos que farão parte da estrutura a ser rastreada por cada negócio.

No Módulo IV, é realizado o registro dos dados relativos a produção para que se tenha a rastreabilidade efetiva da cadeia produtiva. Por fim, no Módulo V, realiza-se as consultas dos dados e a geração de Etiquetas QR-Code (*Quick Response Code*), contendo as informações armazenadas no banco de dados e disponíveis para consulta da empresa e/ou usuário final.

Mantuani (2017) realizou a extensão do RastroGrão com os conceitos de geoinformação. Isso possibilitou que, no Módulo V fossem disponibilizadas informações sobre o local e entorno do local dos produtos a serem rastreados, permitindo que o processo de georastreabilidade fosse integrado durante todas as fases inerentes a rastreabilidade.

3 METODOLOGIA

Nesse Capítulo, na seção 3.1, será apresentado o material e métodos adotados para esta dissertação, relacionando as etapas de desenvolvimento aos objetivos específicos. Na seção 3.2 é realizada a caracterização da pesquisa quanto a abordagem, natureza, objetivos e procedimentos.

3.1 MATERIAL E MÉTODOS

O Quadro 1 apresenta as etapas de desenvolvimento utilizados nesta dissertação.

Quadro 1 – Etapas de desenvolvimento da pesquisa

Objetivo específico	Etapas do Método
Possibilitar que os dados sejam coletados automaticamente	<ul style="list-style-type: none"> • Estudo de arquiteturas de IoT aplicadas na agricultura; • Especificação de uma arquitetura para o <i>framework</i> de rastreabilidade de grãos; • Avaliação da arquitetura proposta.
Contemplar dimensões de qualidade de dados para a tabela de registro de dados	<ul style="list-style-type: none"> • Revisão da qualidade de dados em sistemas de rastreabilidades de produtos agrícolas; • Identificação de dimensões de qualidade em sistemas de rastreabilidade; • Definição de um modelo de qualidade de dados para o <i>framework</i> de rastreabilidade de grãos;

Fonte: A autora

Inicialmente, realizou-se um estudo de arquiteturas de IoT aplicados na agricultura. Considerando as arquiteturas encontradas, o modelo de referência do projeto IoT-A (WALEWSKI, 2013) e o *framework* de rastreabilidade de grãos (VAZ, 2014), definiu-se uma arquitetura em camadas que considerasse a qualidade dos dados.

A arquitetura definida foi avaliada seguindo os requisitos apresentados em Pires et al. (2015): interoperabilidade, descoberta e gerenciamento de dispositivos, interface de alto nível, escalabilidade, gerenciamento de grandes volumes de dados, ciência de contexto, segurança e adaptação dinâmica.

Posteriormente, realizou-se uma revisão da literatura sobre a percepção de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas. Para realizar a busca de artigos foram consultadas as bases de dados *Scielo* (*Scientific Electronic Library Online*), *ACM Digital*

Library, IEEE Xplore Digital Library, ScienceDirect e Google Scholar. A busca e análise dos artigos foram realizadas no período entre novembro de 2018 e março de 2019.

Como suporte na organização dos artigos selecionados, utilizou-se a plataforma Start (*State of the Art through Systematic Review*). Foram incluídos os artigos, teses e dissertações, publicados entre 2010 e 2018, que avaliam a qualidade dos dados na rastreabilidade agrícola/alimentar, escritos em língua portuguesa ou inglesa.

Os descritores utilizados foram: *avaliação e sistema de rastreabilidade, avaliação e framework de rastreabilidade, qualidade e dado e rastreabilidade, qualidade e informação e rastreabilidade, qualidade e dado e agricultura, qualidade e dado e alimento, qualidade e dado e grãos, qualidade e dado e frutas*. A utilização por esses descritores teve como objetivo aumentar a sensibilidade da pesquisa.

Para identificar a percepção de qualidade de dados em estudos de sistemas de rastreabilidade de produtos agrícolas, foram selecionados somente trabalhos que mencionam, ao longo do texto, alguma dimensão de qualidade. Foram consideradas as seguintes definições de qualidade de dados:

- *Acurácia* – refere-se a exatidão do dado ou informação, se está correto e é confiável;
- *Precisão* – é a garantia que as informações e dados estão precisos e de acordo com as necessidades e critérios do cliente;
- *Disponibilidade* – capacidade de disponibilizar os dados no momento em que forem solicitados;
- *Consistência* – é a indicação de que não existe contradição na base de dados. Se um dado é importante para a aplicação e o mesmo está em falta na base de dados, entende-se que falta consistência nessa base;
- *Integridade* – garantia de que os dados são armazenados para garantir precisão e consistência;
- *Interoperabilidade* – capacidade de sistemas trocarem dados e informações de maneira eficiente e eficaz.
- *Usabilidade* – garantia de que os dados contribuem para os resultados esperados;
- *Atualidade* – se os dados são atualizados para atender as necessidades das atividades relacionadas.

Após verificar na literatura como a qualidade de dados é percebida por outros sistemas de rastreabilidade de produtos agrícolas, aplicou-se as Fases de Planejamento e Preparação para obter o modelo de qualidade de dados para o *framework* de rastreabilidade de grãos.

Com suporte da ferramenta BrModelo, elaborou-se o modelo relacional e, a partir dele, foram gerados o modelo físico do banco de dados, as classes e os layouts das telas do sistema. Optou-se por utilizar versão 1.36.32.13 do *framework* de rastreabilidade de grãos (LOPES et al., 2016). Essa versão utiliza o Sistema Gerenciador de Banco de Dados PostgreSQL (POSTGRESQL, 2019), para desenvolver o banco de dados. Com suporte da ferramenta Netbeans (NETBEANS, 2019), foram implementadas as classes do sistema utilizando a linguagem Groovy.

Foram criados testes unitários para cada uma das classes criadas, utilizando o *framework* JUnit. O objetivo dos testes é garantir o funcionamento da aplicação após a alteração de alguma funcionalidade.

3.2 CARACTERIZAÇÃO DA PESQUISA

A metodologia adotada nesta dissertação será classificada quanto a abordagem, a natureza, seus objetivos e seus procedimentos (GERHARDT; SILVEIRA, 2009).

Quanto a abordagem esta pesquisa é qualitativa, pois preocupa-se com a compreensão sobre o que se entende por qualidade de dados em modelos/sistemas de rastreabilidade de produtos agrícolas, para que possa ser especificado requisitos importantes para a qualidade dos dados em modelos de rastreabilidade de grãos.

Quanto a natureza esta pesquisa é aplicada, uma vez que se preocupa com a utilização prática, da qualidade dos dados e da internet das coisas, no problema da rastreabilidade de grãos.

Quanto aos objetivos esta pesquisa é exploratória. Essa modalidade de pesquisa possibilitou a familiaridade com o problema da qualidade de dados em sistemas de rastreabilidade de produtos agrícolas.

Quanto aos procedimentos esta pesquisa é bibliográfica, já que realiza o levantamento de trabalhos já publicados com o objetivo de recolher informações sobre o problema abordado. Com esse tipo de pesquisa, torna-se possível explicar um problema ou apresentar um resultado para uma situação, que será a contribuição para a ciência ou área de atuação.

4 ESPECIFICAÇÃO DE UMA ARQUITETURA DE IOT PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS

O gerenciamento automatizado de processos na cadeia produtiva de grãos, contribui na minimização de custos e de tempo de monitoramento, além de proporcionar segurança aos trabalhadores e precisão para a coleta das informações. Quando um sistema de rastreabilidade utiliza meios de captura de dados automáticos, esse sistema também possui as mesmas qualidades e vantagens da automatização.

Nesse Capítulo é especificado uma arquitetura para a rastreabilidade de grãos baseado em Internet das Coisas. A seção 4.1 apresenta a arquitetura baseada em camadas. Na seção 4.2 discute o Modelo de Domínio, o Modelo de Informação e o Modelo Funcional. As seções 4.3 e 4.4 discutem, respectivamente, os trabalhos correlatos e a avaliação da arquitetura proposta. Por fim, a seção 4.5 realiza as considerações do Capítulo.

4.1 ARQUITETURA BASEADA EM CAMADAS

A arquitetura proposta para o *framework* de Rastreabilidade de grãos é baseada em camadas. Essa arquitetura possibilita que cada camada realize modificações sem interferir nas demais, desde que a interface se mantenha.

A arquitetura contém cinco camadas, a camada de Aplicação, de Dados, de Serviço, de Comunicação e de Dispositivos, além disso, possui dois componentes, Segurança e Qualidade de dados. Salienta-se que o objetivo é realizar a comunicação entre dispositivos da IoT e o *framework* de rastreabilidade de grãos (Figura 6).

Figura 6 – Arquitetura de IoT proposta em camadas

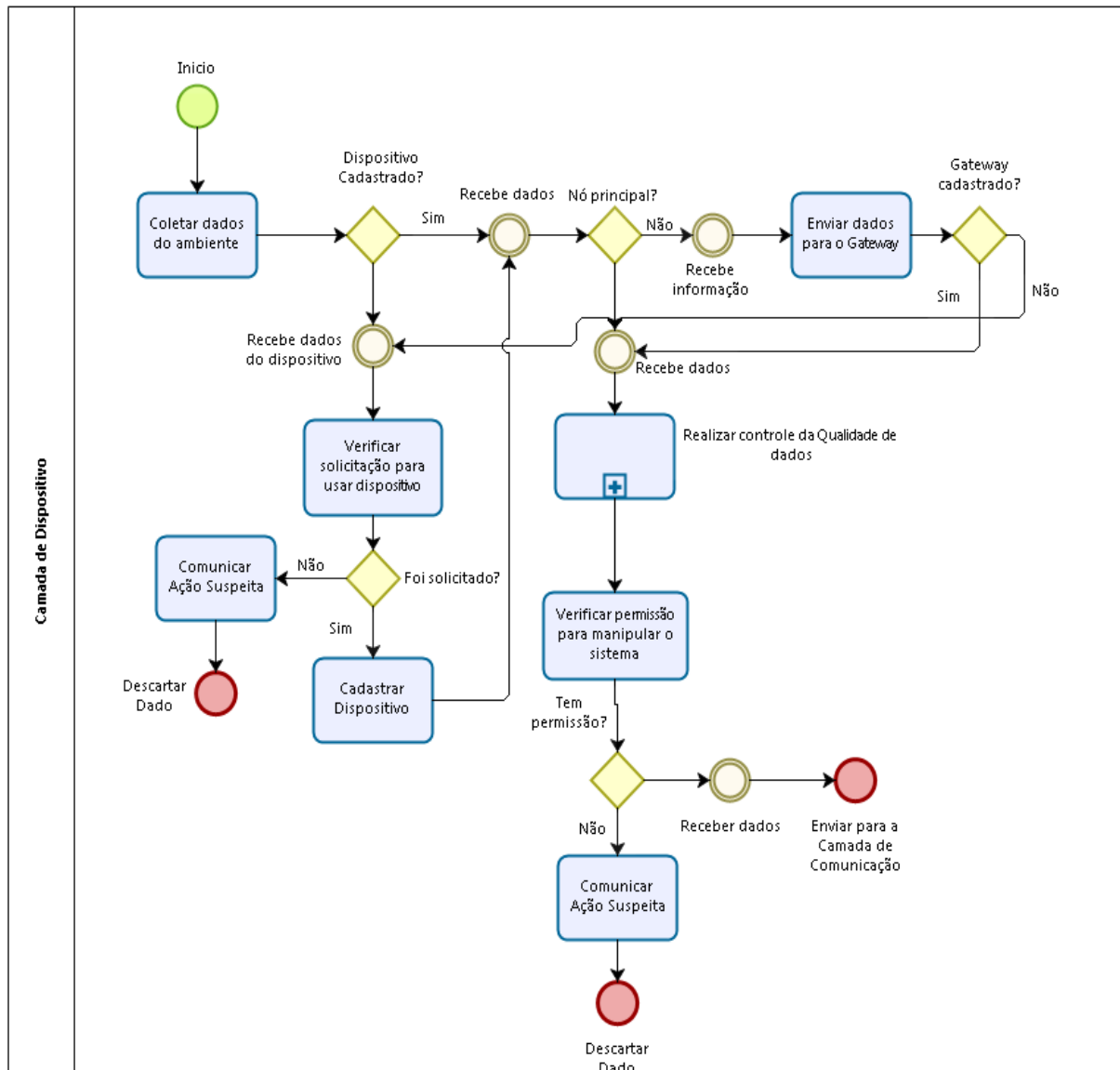


4.1.1 Camada de Dispositivos

A camada de Dispositivos é a camada responsável por obter os dados dos objetos inteligentes ou para interagir com o meio ambiente. São considerados como sendo objetos inteligentes qualquer *hardware* capaz de obter dados como sensores, nós *gateways*, câmeras, veículos aéreos não tripulados, leitoras de etiquetas, etiquetas do tipo RFID e NFC e etc. Cada dispositivo é identificado unicamente no sistema de rastreabilidade e pode, ou não, se comunicar com outros dispositivos por meio de protocolos.

A Figura 7 mostra o fluxo de informação para a obtenção de dados na camada de Dispositivos.

Figura 7 – Fluxo de informação na camada de dispositivos para obtenção de dados



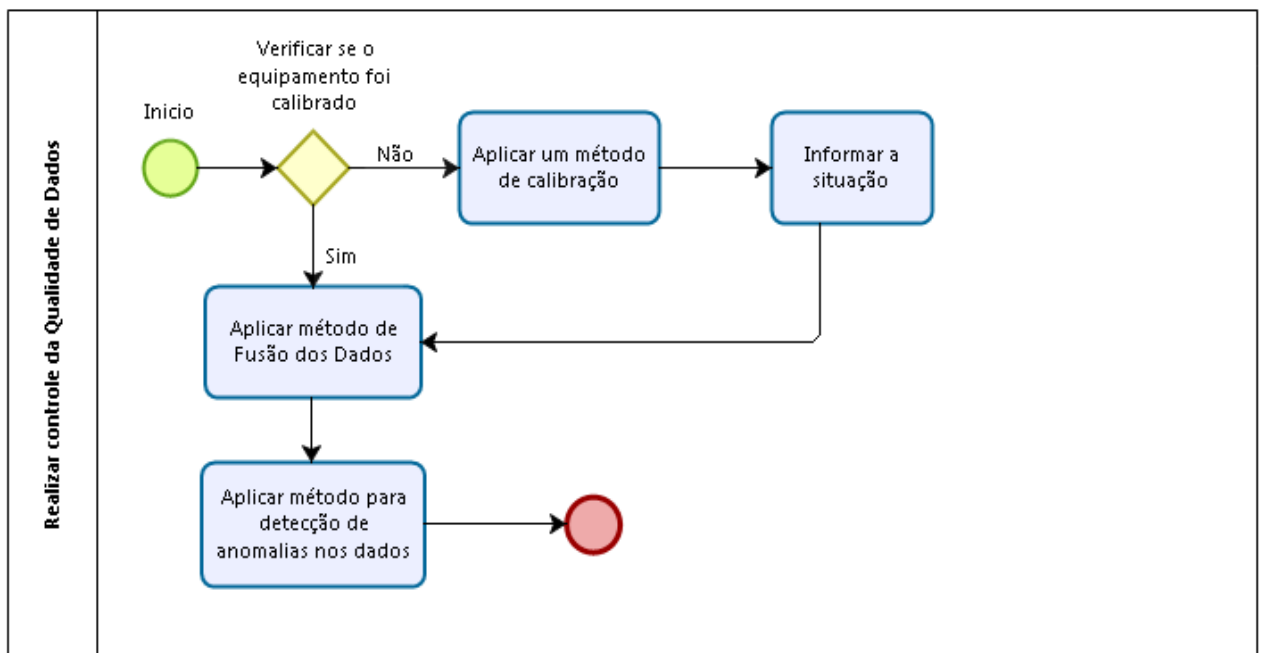
Na atividade “Coletar dados do ambiente”, além da obtenção dos dados desejados como, por exemplo, dados de temperatura e umidade, devem ser coletados dados que descrevem o dispositivo. Manter o histórico da localização e, em casos de redes de sensores, os seus vizinhos, possibilita que sejam realizadas ações de segurança.

Sugere-se a realização de atividades de verificação de solicitação de uso de dispositivo para que, em casos de utilização de redes de sensores, por exemplo, a integridade dos dispositivos se mantenha.

Além disso, antes de enviar os dados coletados para a camada de Comunicação, recomenda-se realizar atividades de qualidade de dados como, por exemplo, fusão de dados e métodos para a detecção de dados anômalos.

Essa qualidade de dados especificada na camada de dispositivos, tem por objetivo melhorar os dados que serão transferidos para o sistema de rastreabilidade. A Figura 8, apresenta o fluxo de informação sugerido para a atividade de “Realizar o controle de Qualidade de dados” na camada de Dispositivo.

Figura 8 – Fluxo de informação do subprocesso “Realizar controle da Qualidade de dados”



Fonte: A Autora

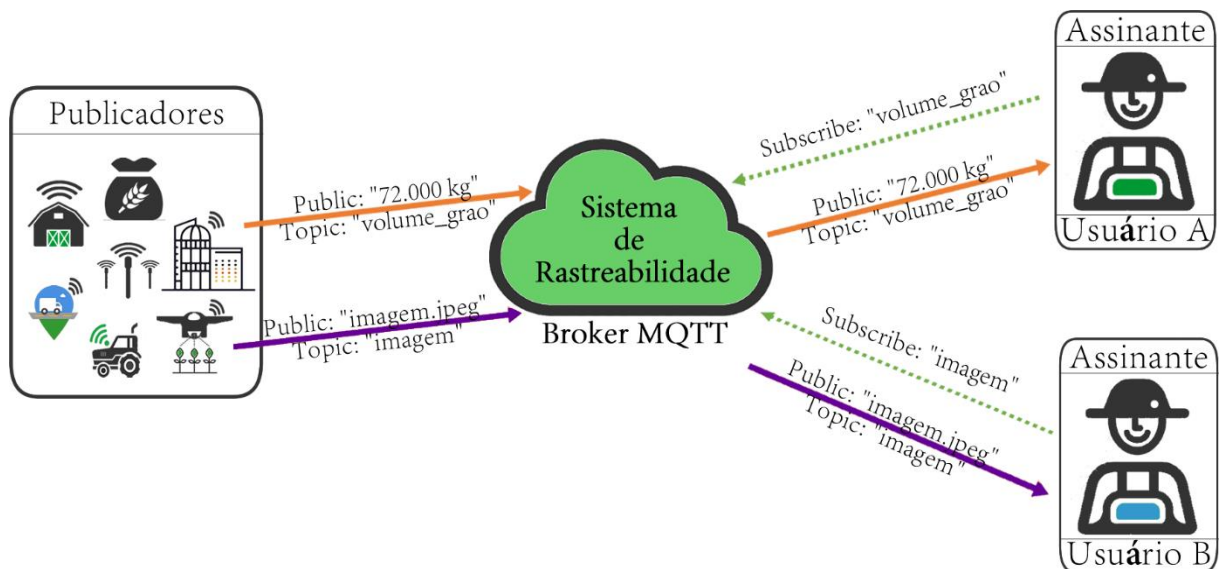
4.1.2 Camada de Comunicação

A camada de Comunicação é responsável por prover mecanismos de comunicação entre o sistema de rastreabilidade e os dispositivos da rede. Assim, antes de realizar a

persistência de um dado no sistema, deve-se verificar se o canal de comunicação, entre os dispositivos e o *framework*, está ativo. Essa ação envolve o relacionamento com outras camadas como a de Dados e a de Serviço.

Na arquitetura proposta, apresentada pela Figura 9, as entidades físicas estão ligadas fisicamente a dispositivos microcontrolados, enquanto que esses dispositivos se conectam ao *framework* de rastreabilidade utilizando conectividade sem fio.

Figura 9 – Troca de mensagens entre Publicadores, Broker e Assinantes



Fonte: A Autora

A troca de mensagens, entre os dispositivos e o *framework* de rastreabilidade de grãos, é realizada por meio do protocolo MQTT. Como o protocolo MQTT utiliza o paradigma *publish/subscribe*, onde as mensagens enviadas pelos publicadores são processadas em um tópico específico, o próprio *framework* de rastreabilidade funcionará como um *broker*, responsável por receber e gerenciar as mensagens que serão enviadas.

Nessa perspectiva, os microcontroladores, que capturam dados e/ou informações sobre um fenômeno de interesse, são chamados de publicadores, enquanto que os usuários, cadastrados para manipular e gerenciar os dados que deverão ser rastreados pelo sistema, são chamados de assinantes.

Caso seja detectado alguma falha na comunicação, é necessário que os dispositivos possuam mecanismos para que, assim que possível, realizem o armazenamento dos dados no *framework* de rastreabilidade.

4.1.3 Camada de Serviço

A camada de Serviço é responsável por gerenciar, associar e modificar os dados coletados pelo sistema da IoT. Um serviço pode ser utilizado para controlar atuadores, obter dados e ou/informações fornecidos por um dispositivo ou por um recurso de rede.

Cada agente da cadeia produtiva pode disponibilizar os seus próprios serviços. Portanto, um usuário só poderá ter acesso a um determinado serviço, após a sua autenticação ao sistema e análise do perfil de acesso.

4.1.4 Camada de Dados

A camada de Dados tem por objetivo gerenciar os dados referentes a rastreabilidade de grãos. Essa camada é a responsável por relacionar os dados, coletados pelos dispositivos, com os objetivos do sistema. Antes de inserir ou atualizar um dado, os dados devem ser validados pelo componente de Qualidade de dados.

4.1.5 Camada de Aplicação

Por fim, a camada de Aplicação refere-se ao *framework* de Rastreabilidade de grãos, ou seja, a interface com o usuário. É por meio da camada de Aplicação que o agente gerencia as especificações da rastreabilidade, dos dispositivos, serviços e dados da IoT.

Por tanto, é na camada de aplicação em que ocorre a adaptação do modelo de rastreabilidade de acordo com as necessidades de cada agente da cadeia produtiva de grãos.

4.1.6 Outros Componentes da Arquitetura

Os componentes Segurança e Qualidade de Dados possuem características comuns às cinco camadas da arquitetura.

4.1.6.1 Segurança

A segurança dos dados deve ser mantida em cada uma das camadas. Como a segurança é um requisito importante em um sistema de rastreabilidade, deve-se gerenciar os dispositivos com acesso ao sistema. Assim, somente dispositivos cadastrados no sistema, podem atualizar

um dado/informação sobre um atributo a ser rastreado. O objetivo é manter a integridade do sistema e dos dados armazenados.

Além disso, cada agente da cadeia mantém os seus próprios serviços de IoT. O requisito de segurança, nesse caso, é garantir que os dados de cada serviço estarão disponíveis somente para quem o requisitou.

4.1.6.2 Qualidade de dados

Como a qualidade do sistema de rastreabilidade deve ser mantida, independente do controle de qualidade de dados da camada de dispositivos, o componente de Qualidade de dados tem por objetivo a validação dos dados gerados. O propósito é garantir a avaliação positiva das dimensões de qualidade de dados que serão escolhidos pelo usuário.

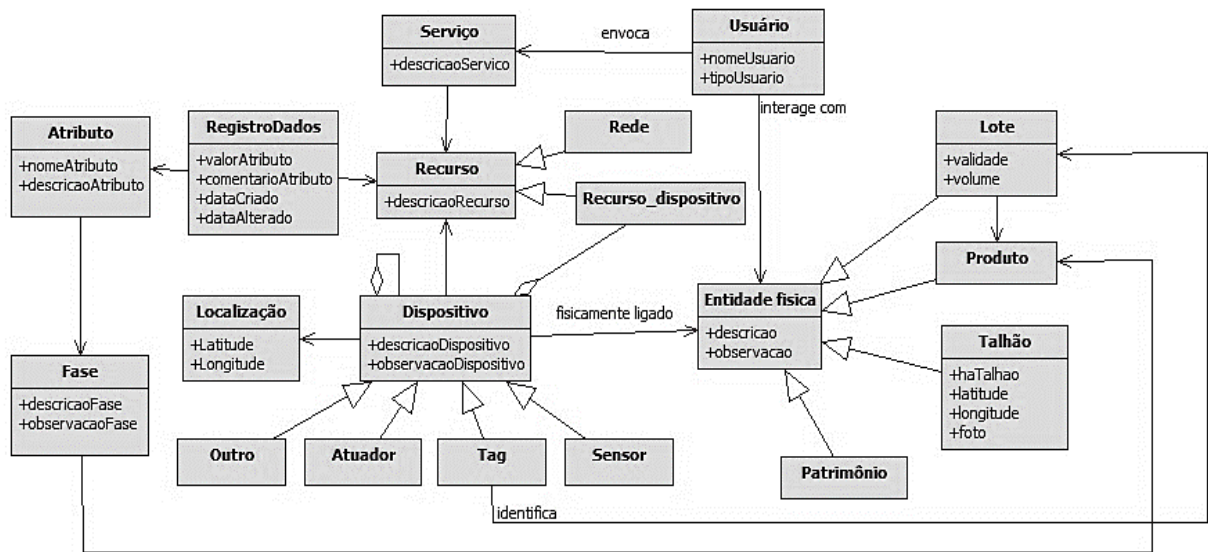
4.2 MAPEAMENTO DA SOLUÇÃO NOS MODELOS PROPOSTOS NA IOT-A

O mapeamento da solução com base na IoT-A, considera o Modelo de Domínio, o Modelo de Informação e o Modelo Funcional. Os modelos serão discutidos, respectivamente em 4.3.1, 4.3.2 e 4.3.3.

4.2.1 Modelo de Domínio

Segundo o projeto IoT-A, a Modelagem de Domínio é o primeiro passo para projetar um sistema para a Internet das Coisas. Considerando as especificações de Vaz (2014), a Modelagem de Domínio proposta (Figura 10), descreve os dispositivos, entidades físicas, recursos e serviços da IoT que pertencem ao cenário de rastreabilidade de grãos, bem como as suas relações.

Figura 10 – Modelo de domínio



Fonte: A Autora

No contexto da cadeia produtiva de grãos, entende-se por Entidades físicas os lotes, produtos, talhões e outros patrimônios de interesse como, por exemplo, caminhões ou silos, que podem ser da propriedade produtora de grãos, ou da empresa gestora dos dados.

Os dispositivos podem ser do tipo Sensores, Atuadores, Tags e Outros. Os sensores são os responsáveis por monitorar o ambiente físico e coletam dados sobre, por exemplo, a temperatura, umidade, imagens e etc.. Já os atuadores modificam o ambiente de acordo com os comandos recebidos por sensores ou pelo sistema de rastreabilidade.

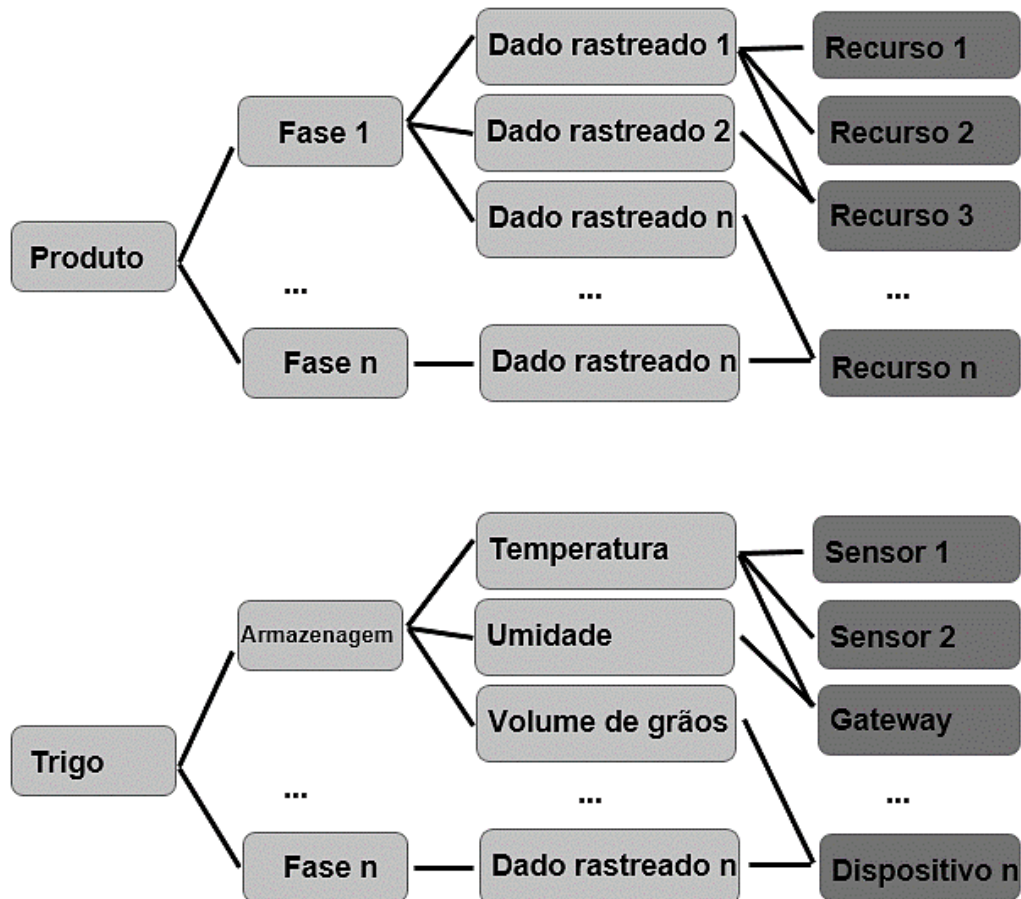
As Tags são identificadores que contêm dados sobre um lote, além de serem um elemento das etiquetas RFID. Dispositivos do tipo Outros, podem ser qualquer *hardware* que deseja manter controle como, por exemplo, veículos aéreos não tripulados, leitoras de etiquetas ou outros servidores de aplicação.

O *framework* de rastreabilidade, hospedado em um servidor *web*, permite que os usuários da aplicação acessem, em tempo real, a dois serviços de IoT. O primeiro é o de monitoramento de uma entidade física como, por exemplo, por meio da utilização de etiquetas de identificação. Já o segundo diz respeito ao gerenciamento do atributo a ser rastreado, durante as fases da cadeia produtiva de grãos.

A gestão da estrutura de rastreabilidade, de qualquer tipo de grão, é realizada por meio do gerenciamento de Produtos, Fases e Atributos (ou Dados) a serem rastreados. Para gerenciar o atributo a ser rastreado durante as fases da cadeia produtiva de grãos, realiza-se a extensão da estrutura de rastreabilidade de Vaz (2014) com a inserção do gerenciamento de Recursos.

Esses Recursos auxiliam na coleta dos dados sobre os Atributos rastreados, seja pela utilização de recursos dos dispositivos, pela utilização dos recursos de outros *softwares* ou pela utilização de funcionalidades do próprio *framework*, como em casos em que ainda seja necessário o preenchimento manual de dados (Figura 11).

Figura 11 – Estrutura de Customização para o *Framework*

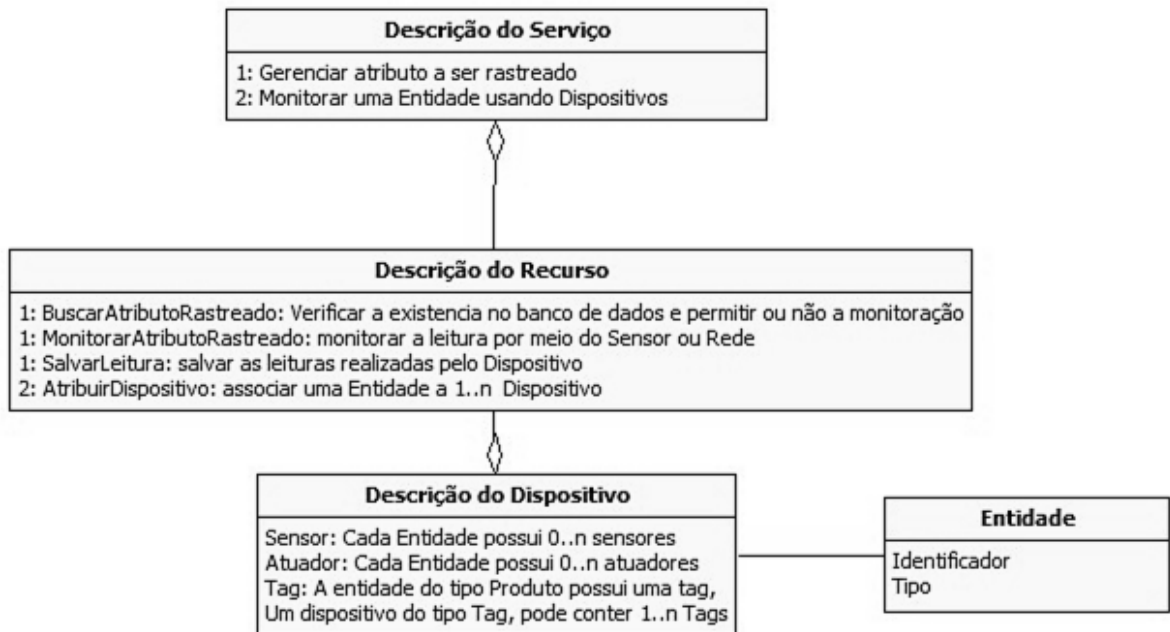


Fonte: A Autora

4.2.2 Modelo de Informação

O Modelo de Informação tem por objetivo modelar as informações pertinentes aos dispositivos, serviços, recursos ou entidades. A Figura 12 representa o modelo de informação, apresentando a descrição do serviço, recurso e dispositivo, bem como o relacionamento entre eles.

Figura 12 – Modelo de informação



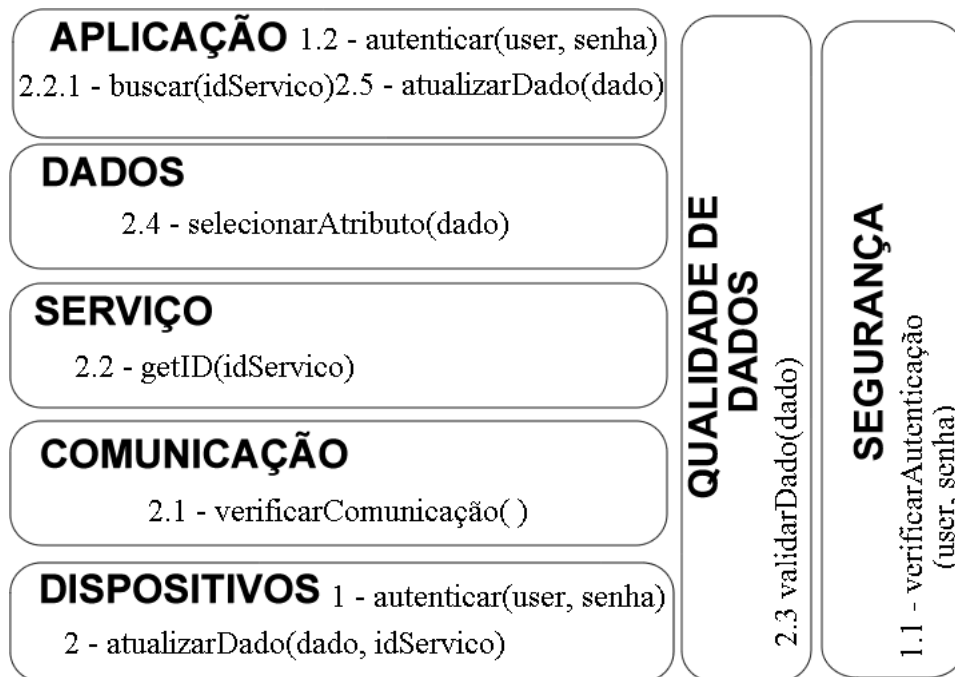
Fonte: A Autora

4.2.3 Modelo Funcional

O Modelo Funcional é o último modelo previsto pela IoT-A. Baseando-se na arquitetura em camadas, apresentado anteriormente, o modelo funcional apresentado tem por objetivo mostrar o relacionamento entre as camadas de Aplicação, de Dados, de Serviço, de Comunicação e de Dispositivos, além dos componentes de Segurança e Qualidade de dados.

Esse modelo, ilustrado pela Figura 13, representa a infraestrutura da IoT, exemplificando a interação entre as camadas para salvar um dado coletado por um dispositivo.

Figura 13 – Modelo funcional



Fonte: A Autora

4.3 TRABALHOS CORRELATOS

Em Krampe et al. (2018) é proposta uma arquitetura híbrida-oportunista para redes agrícolas robustas. A abordagem é de que os pacotes podem ser transferidos com mais eficiência, caso o nó receptor e o remetente se comuniquem diretamente, utilizando a comunicação Máquina para Máquina (M2M). Nos casos em que não seja possível a comunicação M2M, a comunicação via Internet seria o segundo caminho mais rápido. Caso o remetente não tenha conexão com a internet, o pacote deve ser encaminhado baseado em Redes Tolerantes a Atrasos (*Delay Tolerant Networks – DTNs*).

O modelo de Krampe et al. (2018) é composto por três componentes. O primeiro é o componente Sofia-Socket que tem por objetivo enviar e receber pacotes, além de armazenar uma agenda de pacotes e garantir a confiabilidade da comunicação. O segundo componente é o Sofia-Management-Server que armazena os dados. O terceiro componente é o Sofia-Server-Socket que tem por objetivo retransmitir pacotes e fornecer IP (*Internet Protocol*) estáticos.

A arquitetura de Voutos et al. (2018) é aplicada na viticultura. No modelo, os dados são coletados por sistemas eletrônicos e vinculados à uma plataforma *web*, em que a transferência de registros é realizada por meio da comunicação local e remota.

Semelhante ao de Voutos et al. (2018), na arquitetura de Ferrández-Pastor et al. (2018), os dados também são coletados pelos nós sensores e enviados para uma plataforma em nuvem, por meio de um nó *gateway*, que realiza o gerenciamento dos dados e análise de processos. Por meio de uma interface, os usuários podem acessar os dados disponibilizados na nuvem.

Montoya et al. (2017) propõem uma arquitetura de IoT para a agricultura de precisão, com objetivo de coletar, monitorar e analisar variáveis climáticas. A proposta considera as camadas de captura dos dados, a camada de armazenamento, a camada de processamento e a camada de consulta.

A camada de captura tem por objetivo adquirir os dados meteorológicos por meio da utilização de sensores e de um sistema embarcado. A camada de armazenamento recebe os dados e armazena em um sistema online. A camada de processamento, analisa os dados por meio da mineração de dados e realiza previsões e recomendações. Já a camada de consulta monitora as condições de cultivo e fornece a visualização das informações fornecidas pela camada de processamento (MONTROYA et al., 2017).

Em Mekala e Viswanathan (2017) a arquitetura para a agricultura inteligente é baseada em quatro camadas: Camada de aplicação, camada de gerenciamento de informações, camada de gerenciamento de rede e camada de coleta de informações. Os dados coletados sobre o cultivo são enviados para uma plataforma *online* por meio de *gateway*. A partir da nuvem e da análise desses dados, os agricultores podem realizar a tomada de decisão.

Já a arquitetura de Ordoñez-García et al. (2017), aplicada na agronomia de precisão, é composta por três camadas: a camada de aplicação, camada de controle e camada de infraestrutura.

Popovic et al. (2017) propõe uma arquitetura de IoT para agricultura de precisão e monitoramento ecológico. No modelo, os dados coletados pelos sensores são enviados para uma plataforma em nuvem, direta ou indiretamente por meio de um nó *gateway*. A plataforma fornece aos usuários a visualização dos dados por meio de uma interface.

Campos (2017) propôs uma arquitetura de IoT, baseado em camadas, para a rastreabilidade da cadeia de vinhos. A proposta contempla as camadas: Física, Interconexão, de Dados e Camada de Serviço. Além dessas camadas, possuem os componentes: Qualidade de Dados, Segurança e Adaptabilidade.

Diferente das outras arquiteturas, a especificada por Campos (2017) destaca-se pela preocupação com a qualidade de dados e por possibilitar uma arquitetura adaptativa para a cadeia de suprimentos de vinhos.

A arquitetura de IoT proposta por Bing (2017) é baseada em três camadas: camada de percepção, camada de transporte e camada de aplicação. A camada de percepção caracteriza a coleta dos dados pelos sensores. Na camada de transporte foram aplicadas tecnologias de comunicação móvel e *Bluetooth*. Já a camada de aplicação, refere-se a interface que gerencia os dados, coletados pela camada de percepção, e os transforma em informações sobre diagnóstico de doenças baseado em rede neural, além do monitoramento remoto de frutas e vegetais.

Em Channe et al. (2015) é proposto uma arquitetura para a agricultura inteligente baseada em cinco módulos. O primeiro é SensorKit que representa um dispositivo portátil com sensores do solo e ambiente. O segundo módulo é MobileApp responsável por prover uma interface para os usuários. O módulo AgroCloud representa os mecanismos de armazenamento de dados. O quarto módulo está relacionado a mineração de Big Data, análise e criação de conhecimento. O último modelo é Government e AgroBanks UI, uma interface web que captura informações, do Ministério da Cultura da Índia e de bancos agrícolas, relacionadas a esquemas, subsídios e empréstimos agrícolas.

Verdouw et al. (2015) propõem uma arquitetura de referência baseado em IoT para a logística na cadeia de abastecimento agroalimentar. A arquitetura é composta por quatro partes: uma interface local para os usuários interessados na cadeia de fornecimento; um sistema em nuvem para a logística inteligente; um sistema local para logística; e conexões de sistemas externos para gerenciamento da informação do produto.

4.4 AVALIAÇÃO DA ARQUITETURA

A arquitetura pode ser avaliada conforme os requisitos apresentados em Pires et al. (2015). Os requisitos são especificados a seguir.

O requisito *interoperabilidade* está relacionado a capacidade de permitir a utilização de diversos dispositivos e/ou aplicações. Como esses dispositivos podem ser disponibilizados de acordo com as necessidades de negócio, o requisito de *descoberta e gerenciamento de dispositivos*, possibilita que essa integração ocorra dinamicamente e que, com as informações dos dispositivos, permita localizar, apagar, modificar e associá-los a outros dispositivos.

A *interface de alto nível* refere-se à capacidade de adotar uma plataforma de alto nível para acessar os dados e serviços disponibilizados pelos dispositivos físicos. Com isso, a *escalabilidade* também é necessária, já que a integridade do sistema deve ser mantida, mesmo que ocorra requisições intensivas com o crescente número de dispositivos. Além disso, deve-se

realizar o *gerenciamento de grandes volumes de dados*, que se refere a manter os dados que são produzidos pelos dispositivos.

O requisito *ciência de contexto* na IoT, pode ser entendida como as informações adicionais sobre as características do coletor de dados como, por exemplo, a sua localização, estado e, em casos de redes de sensores, os seus vizinhos. O objetivo, nesse caso, é conseguir caracterizar os dispositivos considerando o meio em que estão inseridos.

A *segurança* é outro requisito importante, já que é necessário manter a integridade dos sistemas, dos dispositivos e dos dados envolvidos no processo. Por fim, a *adaptação dinâmica*, está relacionada a disponibilidade e qualidade das aplicações envolvidas, em meio ao dinamismo do ambiente, além da capacidade de adaptação da arquitetura conforme as necessidades do usuário.

No Quadro 2, cada requisito é representado por um número: (1) Interoperabilidade; (2) Descoberta e Gerenciamento de Dispositivos; (3) Interface de Alto Nível; (4) Ciência de Contexto; (5) Escalabilidade; (6) Gerenciamento de Grandes Volumes de Dados; (7) Segurança; e (8) Adaptação Dinâmica. Além disso, o símbolo (*) indica que o requisito é parcialmente atendido e espaços em branco indicam que os requisitos não foram abordados na arquitetura.

Quadro 2 – Avaliação das arquiteturas

Arquitetura	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Krampe et al. (2018)	Sim					Sim	Sim	*
Voutos et al. (2018)			Sim	*		Sim	*	
Ferrández-Pastor (2018)	Sim		Sim			Sim	Sim	Sim
Montoya et al. (2017)	Sim		Sim		Sim	Sim	Sim	
Mekala e Viswanathan (2017)		*	Sim			Sim		
Ordoñez-García et al. (2017)	Sim	*	Sim		Sim	Sim		
Popovic et al. (2017)	Sim	*	Sim			Sim	Sim	
Campos (2017)	Sim	Sim	Sim	*	Sim	Sim	Sim	Sim
Bing (2017)	Sim		Sim			Sim	Sim	
Channe et al. (2015)	Sim		Sim			Sim		
Verdouw et al. (2015)	Sim		Sim		Sim	Sim	Sim	
Arquitetura proposta	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Fonte: A Autora

A arquitetura de Campos (2017) é que a mais se assemelha com a arquitetura proposta. Entretanto, em Campos (2017), o requisito *Ciência de Contexto* não é completamente atendido. Além disso, não é especificado uma camada de Aplicação. O modelo considera que cada agente possui o seu sistema e que são adicionados recursos e serviços.

Outra diferença na arquitetura proposta, com relação a de Campos (2017), é que nessa dissertação o requisito *Adaptabilidade* é proposta na camada de Aplicação, já que cada agente da cadeia produtiva modifica a sua estrutura de rastreabilidade no próprio *framework* de rastreabilidade de grãos. Em Campos (2017), essa Adaptabilidade é proposta em forma de um componente da arquitetura.

4.5 CONSIDERAÇÕES FINAIS

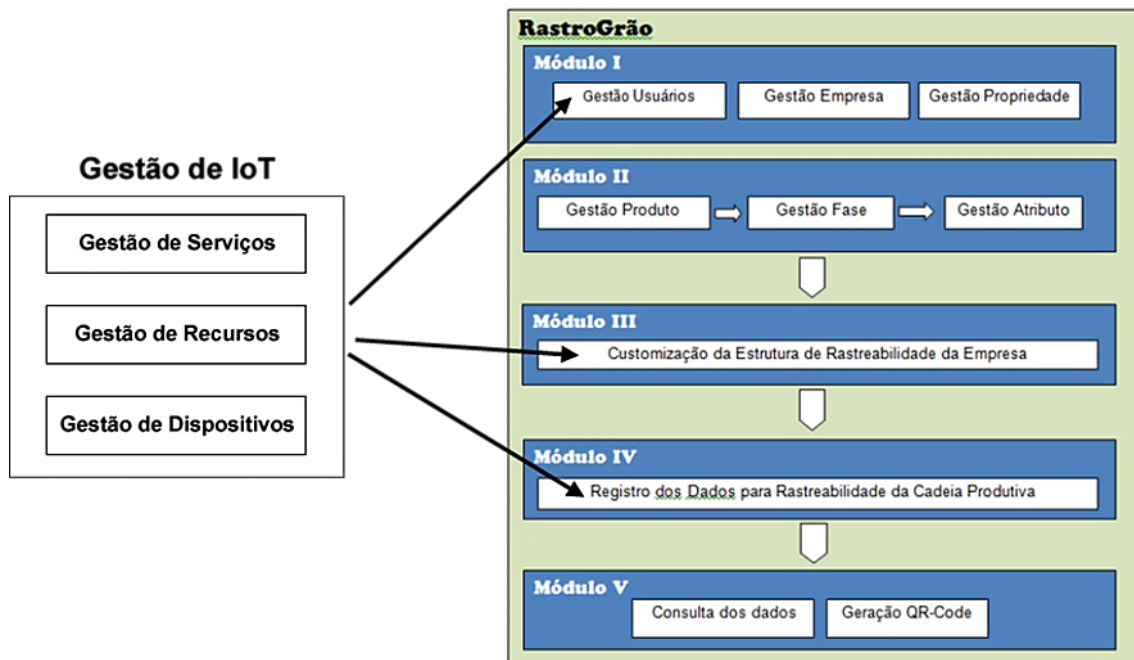
Este Capítulo apresentou uma arquitetura, em camadas e baseada no paradigma *publish/subscribe*, que engloba a Internet das Coisas para a rastreabilidade de grãos. Por meio da customização do processo é possível atender as exigências de cada elo na cadeia produtiva de grãos.

Além disso, apresentou o mapeamento da solução conforme os modelos propostos na IoT-A, considerando o Modelo de Domínio, o Modelo de Informação e o Modelo Funcional.

A avaliação da arquitetura baseou-se nos trabalhos correlatos e nos requisitos do estudo de Pires et al. (2015): Interoperabilidade, Descoberta e Gerenciamento de Dispositivos, Interface de Alto Nível, Ciência de Contexto, Escalabilidade, Gerenciamento de Grandes Volumes de Dados, Segurança e Adaptação Dinâmica.

A estrutura de rastreabilidade de grãos especificada por Vaz (2014) permite a extensão para a incorporação do paradigma de IoT. A Figura 14 ilustra essa extensão e mostra quais as contribuições para o projeto.

Figura 14 – Extensão da estrutura de rastreabilidade de grãos para a integração da IoT



Fonte: A Autora

Em virtude das características dos módulos, no Módulo I são acrescentadas as Gestões relacionadas com a IoT, a Gestão de Serviço, Gestão de Recurso e a Gestão de Dispositivo.

Com os novos dados que passam a ser adquiridas, torna-se possível modificar a estrutura de rastreabilidade da empresa (Módulo III), vinculando os Atributos aos Recursos cadastrados.

O registro dos dados para a rastreabilidade da cadeia produtiva (Módulo IV), no modelo de Vaz (2014), é realizado pelo usuário por meio do preenchimento dos formulários *web*. Posteriormente, em Mantuani (2017), dados geoespaciais podem ser obtidos por meio da integração com bases geoespaciais públicas disponíveis na *web*.

Com a inserção de uma arquitetura de IoT, esse registro dos dados passa a ser também realizada de forma automática por dispositivos, *softwares* e outros recursos que podem ser implementados no *framework* de rastreabilidade.

5 UM MODELO DE QUALIDADE DE DADOS PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS

A qualidade de uma base de dados afeta os resultados de uma análise e podem aumentar o risco de falhas na tomada de decisão. Assim, a QD precisa ser incorporada nos sistemas de rastreabilidade, a fim de garantir que os seus processos são confiáveis. Essa confiança passa pela garantia de que os dados utilizados estão corretos.

Nesse Capítulo, na seção 5.1, são discutidas as percepções de QD em sistemas de rastreabilidade de produtos agrícolas. Na seção 5.2, é definida uma estrutura de qualidade de dados para o *framework* de rastreabilidade de grãos. Na seção 5.3 são identificadas as dimensões de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas. Por fim, a seção 5.4, descreve as considerações finais desse Capítulo.

5.1 QUALIDADE DE DADOS EM SISTEMAS DE RASTREABILIDADE DE PRODUTOS AGRÍCOLAS

O objetivo foi verificar como é a percepção de qualidade dos dados em sistemas de rastreabilidade de produtos agrícolas.

A pesquisa identificou cinco áreas sobre a percepção de qualidade de dados: Coleta automática de dados, Fusão de Dados, Documentação e Modelagem de dados, Especificações de Ontologias e Semântica, além de Modelos de Qualidade de Dados.

5.1.1 Coleta Automática de Dados

A utilização de mecanismos autônomos diminui trabalhos repetitivos e operacionais, facilitam a aquisição de dados, minimizam tempo, custos de gestão e erros como, por exemplo, o preenchimento errôneo de formulários.

Nessa perspectiva, é possível observar a preocupação com os meios de captura dos dados. A automatização de processos, Internet das Coisas, a utilização de tecnologia de identificação por radiofrequência, sensores e/ou Redes de Sensores sem Fio, são formas de proporcionar qualidade aos dados em sistemas de rastreabilidade (TEUCKE et al., 2018; XUEYUAN; BO, 2018; ZENG et al., 2015; SILVA, 2015; EXPÓSITO et al., 2013).

Como os sensores podem monitorar uma variedade de dados como umidade, temperatura e localização, é possível observar trabalhos em que a qualidade dos dados está relacionada com a utilização de sensores como meio de detectar problemas de qualidade de

produtos. Sendo os dados obtidos automaticamente, pode-se realizar medidas de controle de qualidade antes que os incidentes ocorram (TEUCKE et al., 2018; SOMMERFELD et al., 2018).

5.1.2 Fusão de Dados

Os dados de múltiplos sensores possibilitam que sejam obtidas novas informações que, antes, não poderiam ser obtidas utilizando o recurso de somente um sensor. Ao realizar a fusão dos dados desses sensores, também é possível proporcionar qualidade aos dados. O objetivo é obter informações contextualizadas, o que contribui para a tomada de decisão.

Um exemplo da aplicabilidade de fusão de dados é na rastreabilidade de cogumelos. No trabalho de Yao et al. (2018) o objetivo da utilização de fusão de dados foi caracterizar geograficamente amostras de cogumelos. A estratégia adotada foi realizar a fusão dos dados de espectroscopia na região do ultravioleta -visível (UV-Vis) e do infravermelho por transformada de Fourier (FT-IR), combinada com GS-SVM (*Grid Search Support Vector Machine*). Além da estratégia de fusão de dados apresentar bons resultados para a rastreabilidade de cogumelos, o modelo apresenta capacidade de generalização para controlar a qualidade de alimentos semelhantes.

5.1.3 Documentação e Modelagem dos Dados

A modelagem dos dados contribui no entendimento da documentação de processos de negócio e está, diretamente, relacionada com a qualidade da informação extraída do sistema. Por isso, a documentação e a modelagem do banco de dados são outras formas de proporcionar qualidade aos dados.

Para o processo de documentação, a Notação de Modelo de Processos de Negócio (BPMN) pode ser utilizada como ferramenta para análise, captura e abstração lógica dos processos, a fim de extrair conhecimento do cenário estudado, podendo acrescentar dados que serão processados. Além do BPMN, podem ser utilizados o diagrama da Linguagem de Modelagem Unificada (UML) e o diagrama de entidade-relacionamento, com o foco na modelagem dos dados.

A preocupação com a modelagem do banco de dados pode ser observada no processo de produção de suco de frutas, tomate e vegetais (SAVINO et al., 2017), na cadeia de vinhos (VUKATANA et al., 2016), na especificação de um sistema de rastreabilidade de produtos

agrícolas (HAISHUI; LIU, 2010), na rastreabilidade interna em um elevador de grãos (THAKUR et al., 2011), na integração entre sistemas de rastreabilidade de grãos (MING-CHAO et al. 2011), na rastreabilidade da soja (THAKUR; DONNELLY, 2010) e na rastreabilidade de recalls de produtos (LEI; XIAO, 2011; DABBENE; GAY, 2011).

Quando os dados são armazenados corretamente, nos casos em que um lote precisa ser retirado de circulação, a empresa obtém as informações exatas daqueles que estão relacionados com o lote contaminado.

Em Zeng et al. (2015) observa-se a preocupação com a inconformidade e especificação dos dados em sistemas de rastreabilidade de grãos. Para isso, propõem um código de rastreamento, para organizar as informações, com base na Internet das Coisas e níveis de sistemas de negócios. O modelo de rastreabilidade de grãos de Zeng et al. (2015) é baseado em três camadas. Camada de sensor, camada de rede e camada de aplicação.

5.1.4 Especificações de Ontologias e Semântica

A preocupação com a semântica dos dados também é observada como forma de proporcionar qualidade aos dados. Dados fracos semanticamente, causam problemas de consistência de dados e restringem a integração do sistema de origem com outros sistemas de informações.

Nesse sentido, Kaboré et al. (2014) apresenta uma estrutura com padrões e vocabulários comuns para o gerenciamento de dados de trigo, a fim de permitir a interoperabilidade dos dados, facilitando o acesso, descoberta, reutilização e integração dos dados. Já em Deng et al. (2018) é especificado uma ontologia para melhorar a interoperabilidade de dados abertos sobre produtos agrícolas rastreáveis em Taiwan.

Outro exemplo de ontologia é o de Dooley et al. (2018) que tem por objetivo representar entidades que possuem um papel alimentar. A semântica envolve a segurança alimentar, práticas agrícolas e pecuárias ligadas a produção de alimentos, ingredientes e processos culinários, nutricionais e químicos.

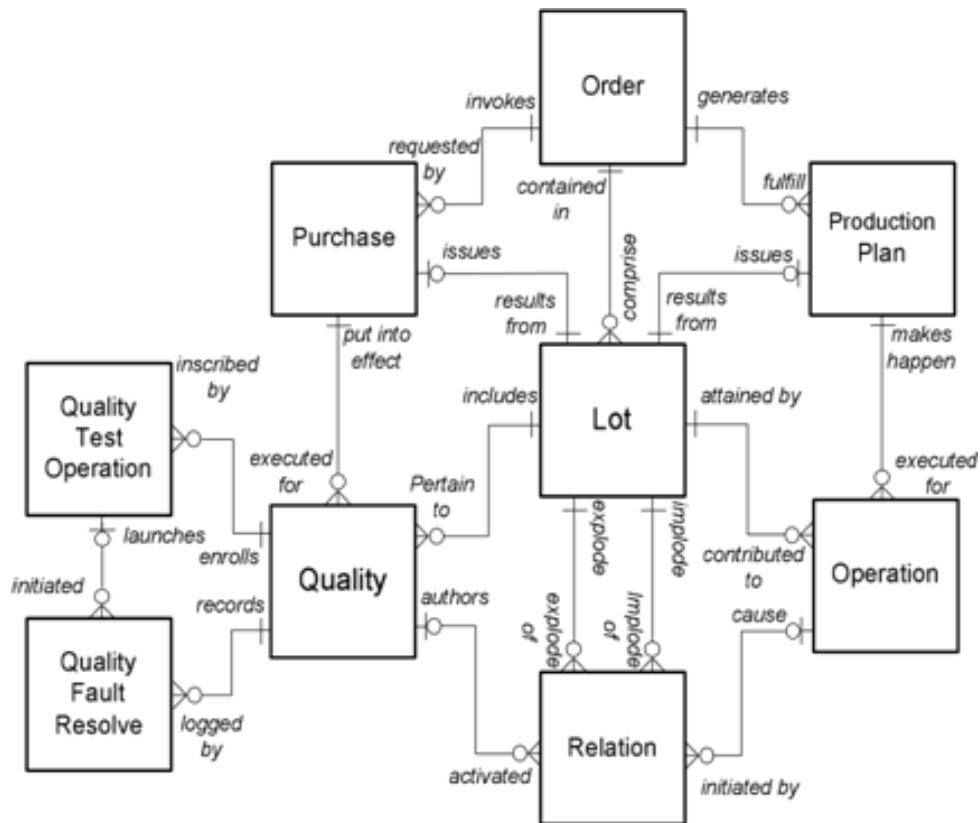
5.1.5 Modelos de Qualidade de Dados

Alguns processos de rastreabilidade como, por exemplo, o de grãos e frutas, podem ser rastreados por meio do gerenciamento de lotes. Nessa perspectiva, o trabalho de Khabbazi et al. (2010) apresenta um modelo para a qualidade de dados em sistemas de rastreabilidade

baseados em lotes. O objetivo é a modelagem do sistema, adequando-o para empresas de pequeno porte com o foco na qualidade da entidade “Lote” em nível de projeto e avaliação.

No modelo, para proporcionar qualidade, são identificadas as entidades Quality (Q), Quality Test Operation (QTO) e Quality Fault Resolve (QFR). A entidade Q identifica as operações de qualidade em um lote e pode incluir testes ou atividades de correção. Já o QTO armazena dados para cada inspeção ou teste de controle de qualidade. Por fim, a entidade QFR é utilizada para armazenar dados sobre as atividades de correção. A Figura 15 apresenta o modelo proposto por Khabbazi et al. (2010).

Figura 15 – Modelo de Qualidade de dados para a entidade “Lot”



Fonte: Khabbazi et al. (2010)

5.2 DEFINIÇÃO DA ESTRUTURA DE QUALIDADE DE DADOS PARA O FRAMEWORK DE RASTREABILIDADE DE GRÃOS

Nas subseções seguintes são descritas as fases realizadas para a definição da estrutura de qualidade de dados para o *framework* de rastreabilidade de grãos.

5.2.1 Fase de Planejamento

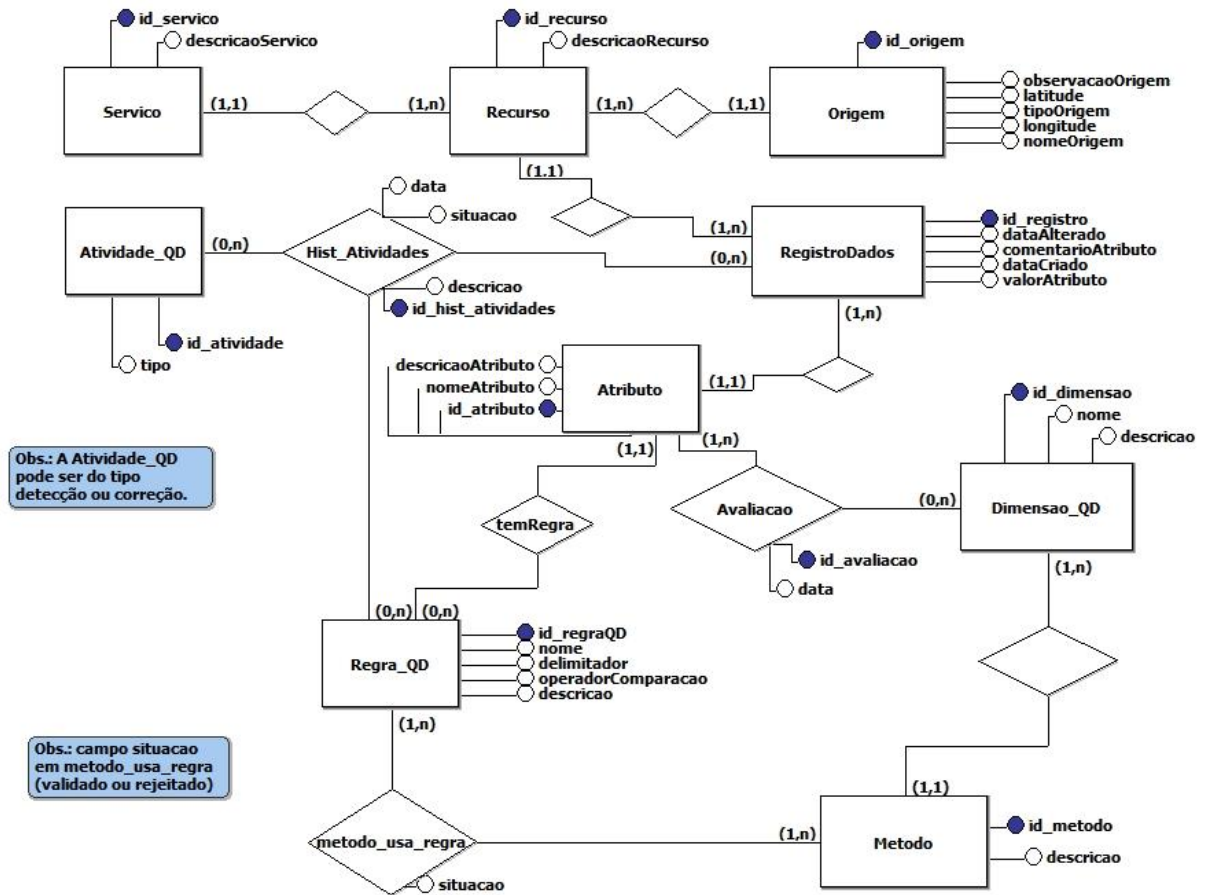
A Qualidade dos Dados pode ser delimitada para um processo específico do negócio. Ao delimitar o estudo é possível saber exatamente quais as tabelas, registros e campos são relevantes.

Nesta dissertação, o escopo de desenvolvimento da qualidade de dados, envolve a tabela de Registro de Dados que, no modelo de rastreabilidade definido por VAZ (2014), tem por objetivo armazenar os dados relativos aos Atributos que são rastreados em cada Fase do processo produtivo de grãos.

Os dados armazenados nessa tabela podem ser utilizados para tomada de decisão. Dessa forma, a qualidade desses dados torna-se uma necessidade para a qualidade do *framework* de rastreabilidade de grãos.

5.2.2 Fase de Preparação

A partir da definição do escopo, foi possível criar o modelo de qualidade de dados. As entidades definidas no modelo de QD são representadas utilizando a modelagem relacional dos dados. A Figura 16 ilustra essa modelagem relacional.

Figura 16 – Modelo de Qualidade de dados para o *Framework* de rastreabilidade de grãos

Fonte: A autora

Uma das formas de proporcionar qualidade aos dados, é por meio da automatização da coleta dos dados. Essa QD é proporcionada pelo fato de que problemas como, por exemplo, preenchimento errôneo de formulários, passam a ser minimizados. Dessa forma, ao possibilitar a coleta dos dados do Atributo por meio das ferramentas da IoT, a qualidade do sistema de rastreabilidade de grãos pode ser melhorada.

Como a qualidade de dados envolve atividades de detecção e correção de problemas, considera-se que a realização constante dessas atividades no banco de dados, colaboram para a qualidade do sistema de rastreabilidade de grãos. Por tanto, neste trabalho, a tabela de Registro de Dados escolhida na definição do escopo, deve passar constantemente por atividades de QD.

Cada atividade de QD deve considerar Regras de QD. Essas Regras de QD estão relacionadas aos Atributos a serem rastreados e envolvem, desde regras para o preenchimento de formulários, até a restrição de valores para o Atributo. Com a criação de Regras de QD para o Atributo, torna-se possível a detecção de anomalias de dados.

Para que os dados do Atributo sejam considerados de qualidade, deve-se o relacionar a uma Dimensão de QD. Ao definir uma Dimensão de QD, é possível associá-lo a um Método de avaliação. Dessa forma, torna-se possível mensurar a qualidade do Atributo.

5.3 IDENTIFICAÇÃO DE DIMENSÕES DE QUALIDADE DE DADOS

Observou-se que, com exceção ao trabalho de Yao et al. (2018), nenhum dos trabalhos selecionados pela revisão da literatura apresentou de forma quantitativa as dimensões de qualidade de dados. Por tanto, as características identificadas não consideram uma métrica, método ou regra para mensurar essa qualidade dos dados.

Como em Yao et al. (2018) a qualidade dos dados é proposta por meio da utilização da técnica de fusão de dados, esse trabalho foi o único que mensurou a dimensão de qualidade. Os autores concluíram que ao utilizar o modelo GS-SVM baseado em fusão de dados, foi possível obter uma acurácia de 100,00% na identificação das origens geográficas dos cogumelos.

A dimensão de qualidade acurácia também é observada no trabalho de Xueyuan e Bo (2018), em que a motivação se encontra na análise de produtos agrícolas para que, em casos de contaminação, seja possível localizar rapidamente as ocorrências dos incidentes.

A preocupação com os dados em Dooley et al. (2018), Deng et al. (2018), Kaboré et al. (2014) e Ming-Chao et al. (2011) refere-se à interoperabilidade dos dados. Dooley et al. (2018) ainda acrescenta que a ontologia descreve os alimentos com acurácia. Já em Deng et al. (2018), além da interoperabilidade, busca-se consistência e usabilidade. Em Ming-Chao et al. (2011), existe também a preocupação com a consistência das informações entre sistemas de rastreabilidade de grãos.

Em Vukatana et al. (2016), Silva (2015), Lei e Xiao (2011), Dabbene e Gay (2011), observa-se que o sistema de informação deve registrar os dados dos clientes com precisão. Isso para que, em casos de recalls, as empresas possam encontrar com eficiência os compradores dos seus produtos. Já em Expósito et al. (2013), a precisão refere-se a busca de produtos e controle de estoque. Em Dabbene e Gay (2011), além da precisão, é possível observar que a acurácia dos dados também é um dos fatores que influencia na qualidade do sistema de rastreabilidade.

A preocupação com a integridade dos dados pode ser observada em trabalhos de modelagem de dados como Savino et al (2017), Zeng et al. (2015), Khabbazi et al. (2010) e Haishui e Liu (2010). Em Thakur e Donnelly (2010) e Thakur et al. (2011), uma das

preocupações é com a integridade dos lotes, o que caracteriza a responsabilidade em manter os dados de entrada e saída do elevador de grãos, para que possam recuperar as informações necessárias desses dados registrados.

Como os processos de qualidade e operação são controlados em tempo real, observa-se em Teucke et al (2018), Sommerfeld et al. (2018), Vukatana et al. (2016), Expósito et al. (2013) e Khabbazi et al. (2010) as dimensões de qualidade: disponibilidade e atualidade.

O Quadro 3, apresenta as dimensões de qualidade identificadas nos sistemas de rastreabilidades encontrados na literatura.

Quadro 3 – Dimensões de qualidade de dados identificados em sistemas de rastreabilidade

Trabalho	Dimensões de qualidade							
	Acurácia	Precisão	Disponibilidade	Integridade	Consistência	Interoperabilidade	Usabilidade	Atualidade
Yao et al. (2018)	x							
Deng et al (2018)					x	x	x	
Teucke et al (2018)			x					x
Dooley et al. (2018)	x					x		
Sommerfeld et al. (2018)			x					x
Xueyuan e Bo (2018)	x							
Savino et al. (2017)				x				
Vukatana et al. (2016)		x	x					x
Silva (2015)		x						
Zeng et al. (2015)				x				
Kaboré et al (2014)						x		
Expósito et al. (2013)		x	x					x
Dabbene e Gay (2011)	x	x						
Lei e Xiao (2011)		x						
Ming-Chao et al. (2011)					x	x		
Thakur et al. (2011)				x				
Khabbazi et al. (2010)			x	x				x
Thakur e Donnelly (2010)				x				
Haishui e Liu (2010)				x				
Modelo proposto	x	x	x	x	x	x	x	x

Fonte: A autora

No modelo de qualidade de dados especificado nessa dissertação, cada Atributo a ser rastreado durante as Fases do processo produtivo de grãos, será avaliado por meio de uma dimensão de qualidade de dados. Essa dimensão de qualidade de dados possui um método de avaliação. Essa avaliação trará, ao consumidor dos dados, um valor quantitativo sobre a dimensão de qualidade de dados escolhida pelo usuário.

5.4 CONSIDERAÇÕES FINAIS

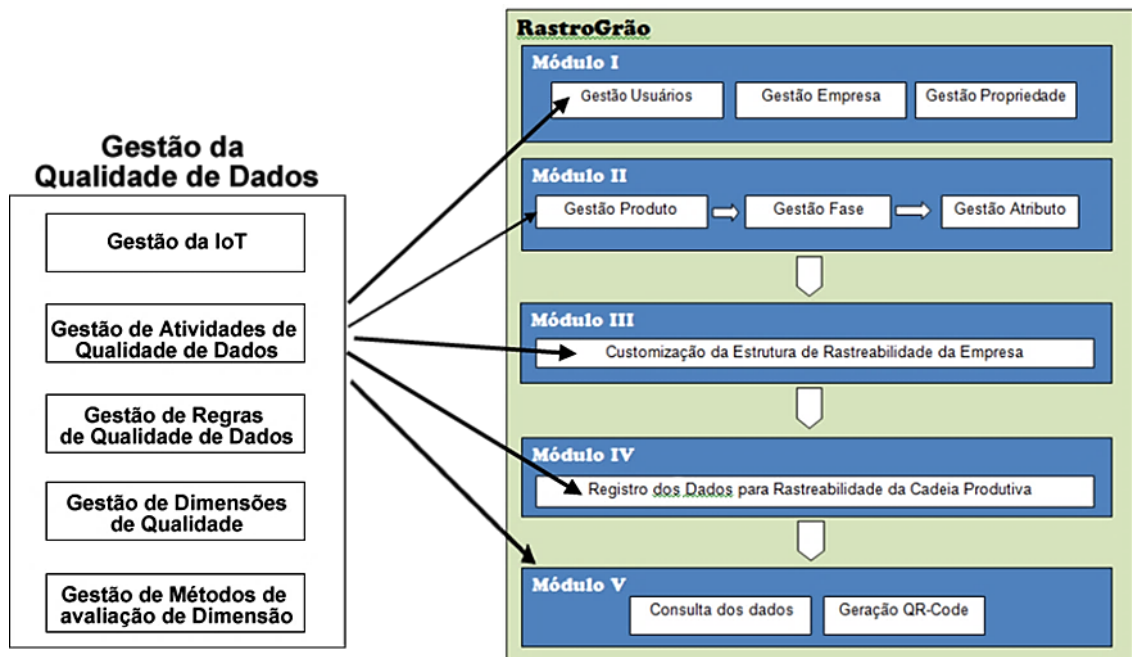
Este Capítulo apresentou a percepção de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas. Com base na revisão da literatura, foi possível observar que a coleta automática dos dados, fusão de dados, documentação e modelagem de dados, especificações de ontologia e semântica, além de modelos de qualidade de dados, são formas de proporcionar qualidade aos dados.

Além disso, foi possível identificar quais as dimensões de qualidade de dados são abordadas nos sistemas de rastreabilidade de produtos agrícolas, sendo elas: acurácia, precisão, disponibilidade, integridade, consistência, interoperabilidade, usabilidade e atualidade.

Observou-se uma carência de avaliação dessas dimensões de qualidade. Nos trabalhos pesquisados, somente o trabalho de Yao et al. (2018), por utilizar uma técnica de fusão de dados, mensurou a dimensão de qualidade.

Por fim, definiu-se uma estrutura de qualidade de dados para o *framework* de rastreabilidade de grãos. A estrutura de rastreabilidade de Vaz (2014), possibilita a sua extensão para incorporar o modelo de qualidade de dados. O modelo de qualidade proposto, baseia-se na tabela de Registro de Dados. Essa tabela refere-se aos Atributos a serem rastreados durante o processo produtivo. A Figura 17 ilustra essa extensão e mostra quais as contribuições para o projeto.

Figura 17 – Extensão da estrutura de rastreabilidade de grãos para a gestão da QD



Fonte: A Autora

A introdução de um modelo de qualidade de dados impacta todos os módulos do *framework* de rastreabilidade de grãos. No Módulo I são acrescentadas a Gestão de IoT, a Gestão de Dimensões de Qualidade e a Gestão de Métodos de avaliação da Dimensão.

No Módulo II, a Gestão do Atributo é estendida para introduzir a Gestão de Regras de Qualidade de Dados, além de introduzir a Gestão de Atividades de Qualidade de dados para a tabela de registro dos dados dos Atributos.

Dessa forma, a estrutura de rastreabilidade da empresa pode ser modificada (Módulo III), vinculando os Atributos aos Recursos cadastradas e as dimensões de qualidade aos seus métodos de avaliação.

Como o registro de dados passa a ser realizada de forma automática, O Módulo IV também sofre alterações. Além disso, ao realizar a Gestão de Qualidade de Dados, os dados consultados (Módulo V) passam a ter uma dimensão de qualidade.

No modelo, a qualidade dos dados é proposta pela: coleta automática dos dados; histórico de atividades de detecção e correção de problemas nos dados armazenados na tabela de registro; além da identificação de dimensões de qualidade para os atributos. Cada dimensão deve possuir um método de avaliação para que seja possível apresentar quantitativamente a qualidade do atributo a ser avaliado.

O resultado da avaliação da qualidade do Atributo, irá auxiliar nas decisões a serem tomadas pelo consumidor dos dados. Essas decisões envolvem a utilização ou não desses dados

armazenados no sistema. Caso a avaliação dos dados seja positiva para o usuário, o mesmo terá certeza de que o processo está correto e poderá prosseguir com a tomada de decisão. Mas caso o resultado mensurado seja negativo, o usuário terá a possibilidade de reverter, pausar ou cancelar o processo, minimizando, por exemplo, custos financeiros e percas de produto.

6 CONCLUSÕES E PESQUISAS FUTURAS

Há a necessidade de obtenção de uma informação correta e em tempo hábil para a tomada de decisão na área agrícola. Para garantir a qualidade da tomada de decisão, deve-se, primeiramente, garantir a qualidade do dado utilizado.

O presente trabalho vem contribuir nas discussões da computação aplicada à agricultura, sobre a qualidade de dados em sistemas aplicados à cadeia produtiva agrícola, por meio da utilização da Internet das Coisas e de um modelo de qualidade de dados.

Nesta dissertação, a coleta automática dos dados tem por objetivo minimizar erros operacionais e melhorar o controle dos dados de um *framework* de rastreabilidade de grãos. Entretanto, não é suficiente, já que os dados coletados ainda podem conter erros.

A obtenção de um modelo de qualidade de dados na rastreabilidade, assim como em outras áreas, beneficia a extração de conhecimento da base de dados e a tomada de decisão. Nesse sentido, este trabalho resulta em um *framework* de rastreabilidade de grãos que, garante a qualidade do Atributo rastreado, por meio da obtenção de dados da IoT, e de critérios de qualidade de dados.

Com relação aos objetivos específicos, o primeiro referente a possibilitar que os dados sejam coletados automaticamente, foi especificado uma arquitetura para o *framework* de rastreabilidade de grãos, originada a partir do estudo de arquiteturas de IoT aplicadas na agricultura. A arquitetura foi avaliada segundo os requisitos apresentados em Pires et al. (2015): Interoperabilidade, Descoberta e Gerenciamento de Dispositivos, Interface de Alto Nível, Ciência de Contexto, Escalabilidade, Gerenciamento de Grandes Volumes de Dados, Segurança e Adaptação Dinâmica.

O segundo objetivo específico foi contemplar dimensões de qualidade de dados para a tabela de registro de dados. Nesse sentido, realizou-se uma revisão da literatura sobre a percepção de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas. Com a pesquisa em bases científicas, identificou as dimensões de qualidade tratadas nesses sistemas e definiu uma estrutura de qualidade de dados para o *framework* de rastreabilidade de grãos, em que fosse possível gerenciar dimensões de qualidade conforme a necessidade do consumidor dos dados.

Dessa forma, como contribuições complementares, destacam-se: (1) Definição de uma arquitetura de IoT para o *framework* de rastreabilidade de grãos; (2) Identificação da percepção de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas; (3) Identificação

de dimensões de qualidade de dados em sistemas de rastreabilidade de produtos agrícolas; (4) Definição de um modelo de QD para o *framework* de rastreabilidade.

Além disso, como resultados do modelo de qualidade de dados, destacam-se: (1) A possibilidade de coleta automática de dados; (2) A realização de atividades de qualidade de dados para a tabela de registro de dados do Atributo rastreado; (3) A definição de regras para o Atributo; e (4) A possibilidade de mensurar a dimensão de qualidade de dados utilizando um método de avaliação.

Na identificação de dimensões de qualidade de dados em modelos de rastreabilidade produtos agrícolas, foi verificado que cada dimensão pode ser aplicada a mais de um contexto como, por exemplo, o caso da dimensão acurácia. Essa dimensão pode ser aplicada para avaliar uma técnica de fusão de dados, mas também é utilizada para se referir à necessidade de localizar rapidamente ocorrências de contaminação.

Não foi encontrado na literatura um modelo de rastreabilidade que considerasse todas as dimensões de qualidade de dados (acurácia, precisão, disponibilidade, integridade, consistência, interoperabilidade, usabilidade, atualidade) simultaneamente. O modelo de qualidade de dados especificado considera, simultaneamente, todas as dimensões e que, por meio de um método possa ser avaliada.

Considera-se fundamental que a qualidade de dados seja introduzida nos modelos e processos de rastreabilidade de produtos agrícolas e, em especial, na rastreabilidade de grãos. A inserção da qualidade de dados nesse contexto contribui para que as tomadas de decisão, realizadas a partir desses dados, também sejam de qualidade.

Como perspectiva de trabalhos futuros sugere-se: 1) Aplicar a arquitetura em um caso real; 2) Simular modelos de RFID e RSSF na rastreabilidade de grãos; 3) Utilizar o modelo de qualidade de dados proposto em outros cenários; 4) Avaliar o modelo de qualidade de dados utilizando dados da IoT; e 5) Especificar métodos para avaliar dimensões de qualidade.

REFERÊNCIAS

- AGGARWAL, C. et al. The internet of things: a survey from the data-centric perspective. In: Aggarwal C. (eds). **Managing and Mining Sensor Data**. Boston: Springer, 2013. p. 383-428.
- AHMED, I.; AZIZ, A. Dynamic Approach for Data Scrubbing Process. **International Journal on Computer Science and Engineering**, v.2, n.2, p. 416-423, 2010.
- ALABA, F. A. et al. Internet of Things security: A survey. **Journal of Network and Computer Applications**. v. 88, p. 10-28, 2017.
- ASHTON, K. **That ‘Internet of Things’ Thing**. 2009. Disponível em: <<http://www.rfidjournal.com/articles/view?4986>>. Acesso em: 14 abr. 2019.
- BADIA-MELIS, R. et al. Food traceability: New trends and recent advances. A review. **Food Control**. v.57, p. 393-401, 2015.
- BATINI, C.; SCANNAPIECO, M. **Data Quality: Concepts, Methodologies and Techniques**. Berlin: Springer, 2006.
- BING, F. The Research of IOT of Agriculture based on Three Layers Architecture. In: 2nd International Conference on Cloud Computing and Internet of Things (CCIOT), 2016. **Anais...** Dalian: IEEE, 2017.
- BORGIA, E. The Internet of Things vision: Key features, applications and open issues. **Journal of Computer Communications**. v. 54, p. 01-31, 2014.
- CAMPOS, L. B. **Uma arquitetura baseada em internet das Coisas para rastreabilidade da cadeia de suprimentos do vinho**. 2017. Tese (Doutorado em Ciências) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2017.
- CHANNE, H. et al. Multidisciplinary Model for Smart Agriculture using Internet-of-Things (IoT), Sensors, Cloud-Computing, Mobile-Computing & BigData Analysis. Hemlata Channe et al. Int. J. **Computer Technology & Applications**. v.6, s/n, p. 374-382, 2015.
- CERUTI, F.C. **Rastreabilidade de grãos: conceito, desenvolvimento de software e estudos de casos de manejo de insetos no armazenamento**. 2007. Tese (Doutorado em Entomologia) – Universidade Federal do Paraná, Curitiba, 2007.
- CERNACH, T. **Um Modelo de Qualidade de Dados em um Data Warehouse**. 2012. Dissertação (Mestrado em Engenharia de Computação) - Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, 2012.
- COOPER, J; JAMES, A. Challenges for database management in the internet of things. **IETE Technical Review**, v.26, n.5, p. 320-329, 2009.
- DABBENE, F.; GAY P. Food traceability systems: Performance evaluation and optimization. **Computers and Electronics in Agriculture**, v.75, n.1, p.139–146, 2011.

DENG, D. et al. Construction and Reuse of Linked Agriculture Data: An Experience of Taiwan Government Open Data. In: ICHISE, R. et al. **Semantic Technology**. Springer, 2018. p. 367-382.

DERRICK, S; DILON, M. **A Guide to Traceability within the Fish Industry**. Publisher: Eurofish, 2004.

DOOLEY, Damion M. et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. **Science of Food**. v.2, n.23, p.10, 2018.

DWIVEDI, A. K.; VYAS, O. P. Wireless Sensor Network: At a Glance. In: Jia-Chin Lin. **Recent Advances in Wireless Communications and Networks**. Taiwan: IntechOpen, 2011. p. 299-326.

EXPÓSITO, I. et al. Efficient traceability solutions in the wine production by RFID and WSN. In: 7th European Conference on Antennas and Propagation. **Anais...** Gothenburg: EUCAP, 2013, p. 3539-3542.

FERRÁNDEZ-PASTOR, F. J. et al. Precision Agriculture Design Method Using a Distributed Computing Architecture on Internet of Things Context. **Sensors**. v.18, n.6, p.1731-1743, 2018.

GALHARDAS, H. et al. Declaratively Cleaning your Data Using AJAX. In: 16èmes Journées Bases de Données Avancées, 2000. **Anais...** Blois: BDA, 2000. p.24-27.

GAZOLLA et al. Rastreabilidade e variabilidade espacial da qualidade fisiológica de sementes soja em campo de produção. **Revista Brasileira de Tecnologia Agropecuária**. v.1, n.1, p. 65-73, 2017.

GERHARDT, T.; SILVEIRA, D. **Métodos de Pesquisa**. Editora da UFRGS, 2009. 120 p.

GIUSTO, D. et al. **The Internet of Things: 20th Tyrrhenian Workshop on Digital Communication**. New York: Springer, 2010.

HAISHUI, J.; LIU, J. Requirement analysis and structural design on an agricultural product quality fast traceability system. In: The 2nd International Conference on Information Science and Engineering, 2010. **Anais...** Hangzhou: IEEE, 2010, p. 2998-3004.

HUSSAIN, F. Internet of Everything. In: HUSSAIN F. **Internet of Things: Building Blocks and Business Models**. SpringerBriefs in Electrical and Computer Engineering. 2017. p. 1-11.

ISO 22005. **Traceability in feed and food chain. General principles and basic requirements for system design and implementation**. Geneva, Switzerland: International Organization for Standardization, 2007.

KABORÉ, E.; YEUMO, D.; MADALLI, D.; KEIZER, J. Opening and linking agricultural research data. **D-Lib Magazine**. v.20, n. 1, p.1-2, 2014.

KHABBAZI, M.R.; HASAN, M.K.; SULAIMAN, R.; MOUSAVI, A. Extending quality data for lot-based traceability system in SME. International Symposium on Information Technology, **Anais...** Kuala Lumpur, 2010, p. 1158-1163.

KHANNA, A.; KAUR, S. Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. **Computers and Electronics in Agriculture**. v.157, n.1, p. 218-231, 2019.

KRAMPE, F. et al. HOAR: A Hybrid-Opportunistic Architecture for Robust Agricultural Networking. In: 2018 IoT Vertica and Topical Summit on Agriculture - Tuscany (IOT Tuscany), 2018. **Anais...** Tuscany: IOT Tuscany, 2018.

LEI, D.; XIAO, S. Building the information management system of defective product recalls. In: IEEE 3rd International Conference on Communication Software and Networks. **Anais...** Xi'an: IEEE, 2011, p. 372-375.

LI, Y. et al. Towards a theoretical framework of strategic decision, supporting capability and information sharing under the context of Internet of Things. **Information Technology and Management**. v.13, n.4, p.205-216, 2012.

LOPES, L. A. et al. Controle de Versão e Alteração no Gerenciamento de Configuração de Software utilizando metodologias ágeis: Aplicação no desenvolvimento de Framework de Rastreabilidade de Grãos. In: CIACA 2016 Proceedings, Conferência Ibero Americana on Computação Aplicada, 2006. **Anais...** Lisboa: IADIS, 2016.

MANTUANI, S. R. **Integração de geoinformação no framework de rastreabilidade de grãos**. 2017. Dissertação (Mestrado em Computação Aplicada) - Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

MEKALA, M. S.; VISWANATHAN, P. A Survey: Smart Agriculture IoT With Cloud Computing. In: International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), 2017. **Anais...** Vellore: IEEE, 2017.

MQTT. **Frequently Asked Questions**. Disponível em: <<http://mqtt.org/faq>>. Acesso em: 01 jun. 2019.

MAYDANCHIK, A. Data quality for Practitioners. In: MAYDANCHIK, A. **Data Quality Assessment**. Bradley Beach: Technics Publications, LLC, 2007. p. 169-309.

MING-CHAO, et al. Research on Multivariate Grain Information Service Platform. In: International Conference on Intelligence Science and Information Engineering. **Anais...** Wuhan, China 2011, p. 530-533.

MONTOYA, E. A. Q. et al. Proposta de uma arquitetura para agricultura de precisão apoiada em IoT. **Revista Ibérica de sistemas de informação e Tecnologias**. v.28, s/n, p.12, 2017.

NAKAGAWA, E. Y. et al. Reference architectures. In: OUSSALAH, M. C. **Software Architecture 1**, 2014. p. 55-82.

NETBEANS. **Download o NetBeans IDE 8.2.** Disponível em: <<https://netbeans.org/downloads/>>. Acesso em: 01 jul. 2019.

NEW VENTAGE PARTNERS. **Big Data Business Impact: Achieving Business Results through Innovation and Disruption: Executive Summary of Findings.** En línea. 2017. Disponível em: <<http://newvantage.com/wp-content/uploads/2017/01/Big-DataExecutive-Survey-2017-Executive-Summary.pdf>>. Acesso em: 20 de ago. 2019.

OLIVEIRA, P. J. M. **Detecção e Correção de Problemas de Qualidade dos Dados: Modelo, Sintaxe e Semântica.** 2008. Tese (Doutorado Informática - Tecnologia da Programação) - Universidade do Minho - Escola de Engenharia, Portugal, 2008.

OLIVEIRA, S. de. **Internet das Coisas com ESP8266, Arduino e Raspberry Pi.** São Paulo: Novatec, 2017. 240 p.

OLSON, J. E., **Data Quality – The Accuracy Dimension.** San Francisco: Morgan Kaufmann Publishers, 2003. 293 p.

ORDOÑEZ-GARCÍA, A. IoT Architecture for Urban Agronomy and Precision Applications. In: IEEE Internatcional Autumn Meeting on Power, Eletronics and Computing (ROPEC 2017), 2017. **Anais...** Ixtapa: IEEE, p.4, 2017.

PIATTINO. **O impacto da má qualidade de dados no resultado de suas ações.** 2017. Disponível em: <<https://piattino.com.br/blog/o-impacto-da-ma-qualidade-de-dados-no-resultado-de-suas-aco-es/>>. Acesso em: 20 de ago. 2019.

PIRES, P. et al. **Plataformas para a Internet das Coisas.** SBRC, Vitória, 2015. 60 p.

POPOVÍČ, T. et al. Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study. **Computers and Eletonics in Agriculture.** v.140, s/n, p. 255-265, 2017.

POSTGRESQL. **Downloads PostgreSQL.** Disponível em: <<http://www.postgresql.org/download>>. Acesso em: 01 jul. 2019.

SAVINO, M.M. et al. A structured approach to implement product traceability for ISO 9001:2015 based on unified modeling language. In: 11th International Conference on Software, Knowledge, Information Management and Applications, 2018. **Anais...** Malabe: IEEE, 2017, p. 1-8.

SCHEER, F. P. Optimizing supply chains using traceability systems. In: SMITH, I.; FURNESS, A. **Improvising traceability in food processing and distribution.** Cambridge, UK: Woolhead publishing. 2006. p. 52-64.

SETHI, P; SARANGI, S. R. Internet of Things: Architectures, Protocols, and Applications. **Journal of Electrical and Computer Engineering.** v.1, p. 1-25, 2017.

SHELBY, Z., et al. **The Constrained Application Protocol (CoAP).** Internet Engineering Task Force (IETF), RFC 7252. 2014.

SILVA, Roberto F. da. **Sistema de rastreabilidade para grânéis sólidos agrícolas: o caso do açúcar**. 2015. Dissertação (Mestrado em Ciências) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção. São Paulo. 2015.

SOMMERFELD, D. et al. Identification of Sensor Requirements for a Quality Data-based Risk Management in Multimodal Supply Chains. **Procedia CIRP**. v.72, s/n, p.563–568, 2018.

TEUCKE, M. et al. Using Sensor-Based Quality Data in Automotive Supply Chains. **Machines**. v. 6 n.4. p.53, 2018.

THAKUR, M.; MARTENS, B.J.; HURBURGH, C.R. Data modeling to facilitate internal traceability at a grain elevator. **Computers and Electronics in Agriculture**. v.75, p. 327–336, 2011.

THAKUR, M.; DONNELLY, K.A.M. Modeling traceability information in soybean value chains. **Journal of Food Engineering**, v. 99, p. 98–105, 2010.

TZOUNIS, A., et al. Internet of Things in Agriculture, Recent Advances and Future Challenges. **Biosystems Engineering**, v. 164, p. 31-48, 2017.

VAZ, M.C.S. **Especificação de um Framework para Rastreabilidade da Cadeia Produtiva de Grãos**. 2014. Dissertação (Mestrado em Computação Aplicada) - Universidade Estadual de Ponta Grossa, Ponta Grossa, 2014.

VERDOUW, C. N. et al. A reference architecture for IoT-based logistic information systems in agri-food supply chains. **Enterprise Information Systems**. v.12, n.7, p. 755-779, 2018.

VOUTOS, Y. et al. An IoT-based insular monitoring for smart viticulture. In: International Conference on Information, Intelligence, Systems and Applications (IISA), 2018. **Anais... IEEE**, p. 1-4, 2018.

VUKATANA, K. et al. Wine Traceability: A Data Model and Prototype in Albanian Context. **Foods**. v.5, n.1, p.11, 2016.

WALEWSKI, J. et al. **Internet-of-Things Architecture IoT-A**. Project Deliverable D1.2 – Initial Architectural Reference Model for IoT, 2013.

XUEYUAN, W.; BO, Y. Research and Design of Traceability System of Agricultural Products. In: 2018 International Conference on Engineering Simulation and Intelligent Control (ESAIC), 2018. **Anais... Changsha: IEEE**, 2018, p. 384-388.

YAO, S. et al. Traceability of Boletaceae mushrooms using data fusion of UV–vis and FT-IR combined with chemometrics methods. **J Sci Food Agric**. v.98, n.6, p.2215–2222, 2018.

ZENG, C. et al. Research of traceability platform of grain quality safety based on grain source tracing code," 2015 International Conference on Logistics, Informatics and Service Sciences (LISS), **Anais...**, Barcelona, 2015, pp. 1-3.