CLÓVIS DOS SANTOS JÚNIOR

**Ontologia Aplicada no Mapeamento de Classe de Dados para a
Melhoria do Intercâmbio de Dados na Cotonicultura no Brasil**

São Paulo
2014

CLÓVIS DOS SANTOS JÚNIOR

**Ontologia Aplicada no Mapeamento de Classe de Dados para a
Melhoria do Intercâmbio de Dados na Cotonicultura no Brasil**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do título de Doutor em Ciências

Orientador: Prof. Dr. André Riyuiti Hirakawa

São Paulo
2014

CLÓVIS DOS SANTOS JÚNIOR

**Ontologia Aplicada no Mapeamento de Classe de Dados para a
Melhoria do Intercâmbio de Dados na Cotonicultura no Brasil**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do título de Doutor em Ciências

Área de Concentração: Engenharia de
Computação

Orientador: Prof. Dr. André Riyuiti Hirakawa

São Paulo
2014

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 27 de maio de 2014.

Assinatura do autor _____

Assinatura do orientador _____

Catálogo-na-publicação

Santos Junior, Clovis dos

**Ontologia aplicada no mapeamento de classe de dados para a melhoria do intercâmbio de dados na cotonicultura no Brasil / C. dos Santos Junior. -- versão corr. -- São Paulo, 2014.
116 p.**

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Sistemas agrícolas 2.Ontologia 3.Metadados 4.Dados (Qualidade; Dimensões) 5.Cotonicultura I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Dedico este trabalho à minha esposa e meus pais pelo apoio e compreensão durante essa longa
jornada.

AGRADECIMENTOS

Ao meu orientador professor Doutor Andre Riyuiti Hirakawa, fico imensamente grato por suas orientações e amizade. O período de convívio foi rico em aprendizado para mim. Meu obrigado aos professores Dr. Antônio Saraiva, Dr. Carlos Cugnasca e professora Cristina Borba, a amizade e ensinamentos de vocês foram muito importantes para a concretização desse trabalho.

Agradeço também aos técnicos administrativos, Sra. Lourdes Keiko, Sr. Edson Souza e Sr. Suzano Bitencourt, sempre prestativos em situações diversas.

Aos colegas do Laboratório de Automação Agrícola, Mara, Marcelo Allan e Michel.

Agradeço a minha esposa, meus pais e demais familiares próximos que sempre acreditaram em mim nesse longo processo.

Finalmente, meu obrigado a equipe do DINTER de Cuiabá-MT por promover uma oportunidade única visando à qualificação dos docentes da UFMT. Também agradeço à Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Quero me desculpar por aqueles que não mencionei, fica meu sincero agradecimento.

Ao sensato basta o necessário.
(Eurípedes/484 a.C - 406 a.C)

Resumo

A gestão da cotonicultura depende de dados com qualidade para a criação de informações que possam auxiliar a tomada de decisões nos processos agrícolas. Parte dessa tarefa depende da legibilidade dos dados gerados, necessários para os processos desse segmento. Com isso, o mapeamento das classes de dados mais relevantes é fundamental para a homogeneização de dados e posterior criação de soluções computacionais. Apesar da clareza da necessidade de recursos de tecnologia da informação para melhorar a gestão na agricultura, há poucos padrões desenvolvidos ou em desenvolvimento para esse fim. Os padrões existentes são específicos ou proprietários, dificultando tanto a homogeneização quanto o compartilhamento de dados. Assim, a proposta da pesquisa foi mapear os processos da produção de fibra de algodão no Brasil, por meio de uma ontologia de domínio, resultando em conhecimento para a criação de um padrão de metadados contribuindo com a melhoria no processo de intercâmbio de dados nesse segmento e também para a melhoria da qualidade de dados na cotonicultura por meio da melhoria nas dimensões de: integridade de dados, completeza e acessibilidade. A validação do padrão de metadados resultante foi realizada por meio de uma prova de conceito verificando as contribuições tanto para o desenvolvimento de software quanto para melhoria da integridade e qualidade de dados na cotonicultura.

Palavras-Chaves: cotonicultura, ontologia, dimensões de dados.

ABSTRACT

The management of cotton production depends on quality data to create helpful information to decision-making in agricultural processes. Some tasks depend on the readability of the data generated and necessary to process this segment. Thus, the mapping of classes most relevant data is crucial for homogenizing data and then creating computational solutions. Despite the clear need for IT resources to improve management in agriculture, there are few standards developed or under development for this purpose. Existing standards are specific or legacy, hindering both homogeneity and data sharing. Thus, the research proposal is to map the cotton fiber production processes in Brazil, using the domain ontology, resulting in knowledge to create a metadata standard contributing to the improvement in the data exchange process in this segment and also improving data quality in cotton production, improving data integrity in dimensions such as completeness and accessibility. The validation of the resulting metadata standard was performed using proof of concept to verify the contributions both for software development and for improving the quality and integrity of data in the cotton industry.

Keywords: cotton production, ontology, data dimensions.

LISTA DE FIGURAS

Figura 1 - Estrutura geral da pesquisa	8
Figura 2 - Visão Geral dos processos na Produção do Algodão	12
Figura 3 - Produção de algodão no Brasil	13
Figura 4 - Árvore de Porfírio	17
Figura 5 - Tipos de ontologias	18
Figura 6 - Fragmento de ontologia de alto-nível para comercialização agrícola	18
Figura 7 - Fragmento de ontologia de domínio cultivo e beneficiamento agrícola	19
Figura 8 - Fragmento de ontologia de tarefas para gerenciamento administrativo	19
Figura 9 - Fragmento de ontologia de aplicação para interoperabilidade de dados	20
Figura 10 - Exemplo de classes e atributos	21
Figura 11 - Representação de ontologia não-lógica	22
Figura 12 - Fragmento de ontologia aplicado à cotonicultura	23
Figura 13 - Visão geral de metadados	25
Figura 14 - Visão geral do fluxo de desenvolvimento da pesquisa	35
Figura 15 - Visão geral do cenário abordado na pesquisa	42
Figura 16 - Formalização de processos na cotonicultura com fluxogramas	44
Figura 17 - Formalização dos processos referentes ao preparo do solo	46
Figura 18 - Formalização dos processos referentes à análise de sementes	47
Figura 19 - Formalização dos processos referentes ao Plantio	48
Figura 20 - Formalização dos processos referentes ao Beneficiamento	49
Figura 21 - Delimitação de domínio para extração de conceitos	50
Figura 22 - Classe de dados referentes ao preparo de solo	53
Figura 23 - Classe de dados referentes à análise de sementes	54
Figura 24 - Classe de dados referentes ao plantio	56
Figura 25 - Classe de dados referentes ao beneficiamento	58
Figura 26 - Classes de dados propostas	59

Figura 27 - Resumo das classes de dados com atributos.....	60
Figura 28 - Inconsistência de dados.....	64
Figura 29 - Fragmento do mapeamento entre modelo ExR e classes de dados	66
Figura 30 - Modelo ExR correspondente à estrutura de <i>tags</i>	67
Figura 31 - Fragmento do modelo ExR.....	67
Figura 32 - Mapa ilustrativo de divisão de quadras em fazendas.....	68
Figura 33 - Estrutura geral da prova de conceito.....	69
Figura 34 - Modelo ExR para cotonicultura no centro-oeste brasileiro.....	70
Figura 35 - Estrutura dos protótipos para a prova de conceito.....	70
Figura 36 - Modelo BPEL para criação de serviços para cotonicultura.....	71
Figura 37 - Interfaces dos protótipos desenvolvidos.....	72
Figura 38 - Aplicação com Arquitetura Proprietária.....	73
Figura 39 - Dados formalizados utilizando o padrão de metadados com <i>XML</i>	88
Figura 40 - Integridade de dados em fragmento de metadados.....	89

LISTA DE GRÁFICOS

Gráfico 1 - Soluções corporativas utilizadas no agronegócio	24
Gráfico 2 - Linguagens para formalização de ontologias	31
Gráfico 3 - Empresas brasileiras	39
Gráfico 4 - Empresas estrangeiras	39
Gráfico 5 - Complexidade ciclomática	74
Gráfico 6 - Quantidade de linhas de código	74
Gráfico 7 - Número de membros	77
Gráfico 8 - Quantidade média de caracteres por linha	77
Gráfico 9 - Quantidade média de parâmetros de entrada e saída	76
Gráfico 10 - Uso de funções	76
Gráfico 11 - Análise das dimensões de intercâmbio de dados	78
Gráfico 12 - Tempo de resposta (ms)	86
Gráfico 13 - Espaço de armazenamento (MB)	86
Gráfico 14 - Resumo de métricas de desenvolvimento de <i>software</i>	94
Gráfico 15 - Resumo de métricas de intercâmbio de dados	94
Gráfico 16 - Resumo de métricas de qualidade de dados	95

LISTA DE QUADROS E TABELAS

Quadro 1 - Características dos tipos de representação de conhecimento.....	16
Quadro 2 - Fragmento de ontologia baseado em lógica.....	21
Quadro 3 - Análise de características dos padrões de metadados.....	33
Quadro 4 - Características dos tipos de representação de conhecimento.....	36
Quadro 5 - Locais de visitas técnicas – Acadêmicas.....	40
Quadro 6 - Locais de visitas técnicas – Certificadoras.....	40
Quadro 7 - Locais de visitas técnicas – Comerciais.....	40
Quadro 8 - Locais de visitas técnicas – Governamentais.....	41
Quadro 9 - Locais de visitas técnicas – Tecnológicas.....	41
Quadro 10. Classes de dados - preparo de solo.....	52
Quadro 11. Classes de dados – análise de sementes.....	53
Quadro 12. Classes de dados – plantio.....	55
Quadro 13 - Classes de dados – beneficiamento.....	57
Quadro 14 - Lista de requisitos para prova de conceito.....	62
Quadro 15 – Dimensão Acessibilidade.....	79
Quadro 16 – Dimensão Completude.....	79
Quadro 17 – Dimensão Conformidade.....	80
Quadro 18 – Dimensão Especificação de Dados.....	80
Quadro 19 – Dimensão Facilidade de Uso e Manutenibilidade.....	81
Quadro 20 – Dimensão Fonte.....	81
Quadro 21 – Dimensão Objetividade.....	81
Quadro 22 – Dimensão Apresentação.....	82
Quadro 23 – Dimensão Transabilidade.....	82
Quadro 24 – Dimensão Validade.....	83
Quadro 25 - Comparativo entre padrão csv e dicionário de dados.....	87

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i>
BPEL	<i>Business Process Execution Language</i>
BPMN	<i>Business Process Modeling Notation</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico Brasileiro
CONAB	Companhia Nacional de Abastecimento
CSV	<i>Comma-Separated Values</i>
EDI	<i>Electronic Data Interchange</i>
ERP	<i>Enterprise Resource Planning</i>
HVI	<i>High Volume Instrument</i>
KDD	<i>Knowledge-Discovery in Databases</i>
PoC	<i>Proof of Concept</i>
RH	Recursos Humanos
SI	Sistemas de Informação
SPED	Sistema Público de Escrituração Digital
Tags	Palavra ou termo relevante para representação de dados em linguagens de marcação.
WSDL	<i>Web Services Description Language</i>
UML	<i>Unified Modeling Language</i>
XML	<i>eXtensible Markup Language</i>
XML Schema	Linguagem de marcação baseada em XML para definir regras de validação em documentos no formato XML
Zbig	Z39.50 <i>Biology Implementers Group</i>

SUMÁRIO

1. INTRODUÇÃO	6
1.1 Contexto	6
1.2 Justificativa	6
1.3 Objetivo	7
1.4 Estrutura da Tese	7
2. AGRICULTURA NO BRASIL	9
2.1 Agronegócio	9
2.2 Aspectos Sociais	10
2.3 Cotonicultura	11
2.3.1 Cotonicultura no Brasil	12
2.3.2 Análise de Solos	14
2.3.3 Seleção de Variedades	14
2.3.4 Acompanhamento do Plantio	14
3. ONTOLOGIAS	16
3.1 Classificação	17
3.2 Classes e Atributos	20
3.3 Taxonomia	21
3.4 Intercâmbio de Dados	23
4. METADADOS	25
4.1 Padrões de Metadados na Agricultura	26
4.1.1 <i>Dublin Core</i>	26
4.1.2 <i>Darwin Core</i>	27
4.1.3 <i>AgMES</i>	27
4.1.4 <i>AGRIS</i>	28

4.1.5 Agrovoc	28
4.1.6 AGXML	29
4.1.7 AGROXML	30
4.2 Padrão para Cotonicultura	31
4.3 Dimensões de Qualidade de Metadados	32
4.4 Contribuições Multidisciplinares	33
5. MATERIAIS E MÉTODOS	36
5.1 Delimitação das Classes de Dados	41
5.2 Escopo de Classes abordadas	44
6. MAPEAMENTO DOS PROCESSOS	45
6.1 Detalhamento dos Processos	45
6.2 Mapeamento das Classes de Dados	50
6.3 Padrão de Metadados Proposto	59
7. PROVA DE CONCEITO	62
7.1 Qualidade de Dados Aplicada à Agricultura	62
7.2 Detalhamento da Prova de Conceito	64
7.2.1 Escopo da Prova de Conceito	65
7.2.2 Ambiente para Testes	65
7.2.3 Mapeamento entre Modelos de Dados	66
7.2.4 Infraestrutura	69
7.2.5 Análise dos resultados	73
7.2.5.1 Métricas de Desenvolvimento de Software	73
7.2.5.2 Métricas de Intercâmbio de Dados	77
7.2.5.3 Métricas de Qualidade de Dados	87
7.3. Discussão dos Resultados	90
8. CONCLUSÃO	96
8.1 Trabalhos Futuros	97

9. REFERÊNCIAS	99
-----------------------	-----------

1. INTRODUÇÃO

De acordo com o Ministério da Agricultura, Pecuária e Abastecimento do Brasil (MAPA, 2010), a cultura do algodão no Brasil vem crescendo, especialmente nos Estados de Mato Grosso, Mato Grosso do Sul e Bahia. A exportação de fibra de algodão aumentou duas vezes em apenas uma safra. Em 2003/2004, o Brasil produziu 1,2 milhão de toneladas de fibra, majorando em 46,3% a safra anterior. Segundo a Companhia Nacional de Abastecimento – CONAB, a safra 2011/2012 resultou em uma produção de 1,44 milhão de toneladas, apesar da redução de produtividade na região Centro-Oeste, principal produtora de algodão no país.

1.1 Contexto

A partir da década de 80, soluções computacionais começaram a ser desenvolvidas para a agricultura no Brasil, sempre com a perspectiva de atender a problemas pontuais, como gerenciamento de produção, contabilidade, gestão financeira e gerenciamento de recursos humanos. O incremento de soluções para os problemas supracitados segue diferentes parâmetros, resultando em estruturas para o armazenamento de dados heterogêneos. A ausência de dados homogêneos cria lacunas no compartilhamento e integração de sistemas gerenciais. Grande parte do problema está na verticalização dessas soluções. Atualmente, a integração de bases de dados em aplicações como *Business Intelligence (BI)*, *Data Warehouse* e *Enterprise Resource Planning (ERP)*, Santana (2008) ilustra esse tipo de problema.

1.2 Justificativa

Parte do problema discutido na contextualização desta pesquisa está na ausência de um padrão de intercâmbio de dados no segmento agrícola. Nesse sentido, os metadados podem ser uma importante ferramenta para a criação de mecanismos na permuta de dados. Entretanto, o uso de padrões genéricos de metadados dificulta o intercâmbio de dados entre processos da produção do algodão. Diferentes processos resultam, assim, em estruturas de dados heterogêneos.

Essa demanda justifica a proposta de criação de um padrão para o segmento agrícola, específico para a produção de fibra de algodão. O padrão é um subproduto da formalização abordada na pesquisa de uma ontologia para representar o conhecimento sobre esse segmento, por meio da identificação das principais classes de dados. O benefício direto está na contribuição para atender a demandas na criação de recursos que facilitem a troca de dados.

1.3 Objetivo

O objetivo da pesquisa é contribuir com a integração e a qualidade de dados na cotonicultura, utilizando uma ontologia para estabelecer o conhecimento referente aos processos desse segmento agrícola, formalizando-a em um padrão de metadados, a fim de possibilitar, de forma prática, o seu uso, contribuindo, assim, para a redução do isolamento de sistemas computacionais com estruturas de dados heterogêneos e mapeamento de fluxos de dados para gestores agrícolas. Isso proporcionará a criação de soluções para o acompanhamento, rastreabilidade e atribuição de responsabilidades entre os processos.

1.4 Estrutura da Tese

A pesquisa é estruturada em oito itens:

- 1) Abordagem histórica, social e econômica da agricultura, com ênfase na cotonicultura no Brasil com destaque para o Estado de Mato Grosso.
- 2) Apresentação teórica do conceito de ontologia, com classificação dos tipos, de acordo com áreas de aplicação.
- 3) Enfoque ao conceito de metadados, tipos e aplicações na área agrícola.
- 4) Exposição de conceitos a respeito de qualidade de dados e problemas relacionados às deficiências.
- 5) Contribuição proveniente das formalizações da ontologia e do padrão de metadados.
- 6) Demonstração do padrão de metadados, resultante das classes de dados mais relevantes identificadas e da estrutura de *tags*.

7) Análise dos resultados obtidos por meio da prova de conceito em protótipos implementados para a verificação da melhoria do intercâmbio de dados com o padrão de metadados proposto.

8) Apresentação de melhorias complementares na qualidade de dados, resultantes do aperfeiçoamento no intercâmbio de dados.

A Figura 1 ilustra uma visão geral da estrutura da pesquisa, relacionando contribuições e aplicações para a ontologia e o padrão de metadados resultante; inicia-se com o mapeamento dos dados relacionados à cotonicultura, construção da ontologia, formalização da ontologia em um padrão de metadados e, posteriormente, criação dos protótipos para validação do padrão. Também são apresentados resultados referentes à contribuição do padrão de metadados na melhoria da qualidade de dados, através da homogeneização de dados heterogêneos.

Figura 1 - Estrutura geral da pesquisa.

2. AGRICULTURA NO BRASIL

Segundo Corrêa (2003), “As frentes de agricultura empresarial só começam a atingir o Centro-Oeste no final da década de 1950. Antes disso, no pós-guerra, a agricultura dinâmica do país concentrava-se no centro-sul, em São Paulo, Minas Gerais e Rio Grande do Sul, sendo o norte do Paraná, partes do Triângulo Mineiro, sul de Goiás e Mato Grosso as fronteiras agrícolas a serem descobertas pela agricultura comercial. Isto ocorreu somente na segunda metade da década de 50, na medida em que os problemas de falta de infraestrutura básica e de transporte foram sendo resolvidos [...]”.

Apesar dessa deficiência de infraestrutura ter ocorrido no século passado, considera-se um acontecimento recente. É importante destacar que a produção agrícola da região Centro-Oeste concentrou-se no Estado de Goiás. Mais adiante, na década de 1990, é possível perceber que a retomada da produção nacional, após anos em crise, dá-se quando a cultura do algodão se moderniza e migra para o Centro-Oeste, especialmente para o Estado de Mato Grosso. Prova disso são os crescentes aumentos verificados na produção desse Estado a partir de 1997. Com isso, pode-se dizer que a produção do Centro-Oeste, especificamente em Mato Grosso, foi responsável pelas recentes mudanças ocorridas no mercado nacional de algodão. Essa é, sem dúvida, a cultura com o maior grau de investimento em tecnologia (CORRÊA, 2003).

2.1 Agronegócio

A inovação tecnológica tem contribuído para a melhoria em termos de quantidade e qualidade dos alimentos em todo o mundo. A aplicação da ciência e tecnologia é cada vez mais intensa na totalidade do agronegócio e isso pode ser observado na aplicação e gerenciamento de insumos, produção agrícola e, até mesmo, na pecuária, envolvendo transformação, comercialização e gestão de pessoas e finanças. Os desafios para a agricultura no futuro, no entanto, crescerão ainda mais. Estima-se que a agricultura mundial terá que aumentar a produção de alimentos de forma crescente para abastecer cerca de dois bilhões de novas pessoas (WEICK, 2001).

O grande desafio do homem, no novo milênio, é assegurar as provisões de alimentos e o desenvolvimento sustentável da agricultura. Com uma população mun-

dial em torno de seis bilhões de habitantes, cujo suprimento de alimentos provém de uma área cultivada de 1,5 bilhões de hectares, a produção ainda é insuficiente para atender a demanda, pois se estima que existam 30% da população passando fome ou subnutrida. Segundo as Nações Unidas, o crescimento populacional para 2025 é avaliado em oito bilhões; todavia, a expansão da área cultivada não lhe é proporcional, totalizando algo em torno de 1,6 bilhões de hectares. Assim, a proposta do agronegócio é apresentar formas de desenvolvimento da agricultura, utilizando a tecnologia para melhorar a eficiência nos processos envolvidos, desde o plantio até o beneficiamento de produtos. Esperam-se, com isso, resultados consideráveis, como economia de tempo e trabalho. Exemplo disso pode ser observado na aplicação correta de insumos de forma customizada, o que reduz custos de mão de obra, insumos e equipamentos.

2.2 Aspectos Sociais

Apesar do estabelecimento e consolidação da agricultura no cenário atual, o caminho percorrido desde a década de 1940 foi caracterizado por diversos problemas sociais. Ajudado pelos subsídios governamentais, o Estado de Mato Grosso, foco da pesquisa, evoluiu por meio de pesquisas agrícolas intensas para superar a baixa fertilidade de solos característicos em regiões de cerrado. Assim, a produção agrícola expandiu do sul do país até o vasto cerrado brasileiro. Com o propósito de fortalecer as atividades agrícolas, os grandes agricultores incentivaram pequenos produtores a vender suas terras e então avançar para novas fronteiras em regiões antes compostas por florestas nativas. Essa dinâmica também favoreceu o deslocamento de pequenos agricultores para áreas urbanas, agravando problemas sociais, como o desemprego, criminalidade e segregação. No caso específico do Estado de Mato Grosso, o processo começa com a conversão da floresta, inicialmente em pequena escala, para a extração de madeira, mineração e agricultura de subsistência. A consolidação do uso da terra, em grande escala, para pecuária foi o próximo passo nesse processo expansivo, que teve efeitos importantes, tanto sociais quanto ambientais na região (FERNANDES, 2009).

2.3 Cotonicultura

O escopo da pesquisa é delimitado à cultura do algodão com ênfase na produção de fibra. O algodoeiro, planta da família das malváceas, é conhecido desde 3000 anos a.C., sendo que há 1500 anos a.C. já era cultivado na Índia para a fabricação de tecidos. Mil anos depois, a China produzia tecidos de algodão, produtos somente comercializados na Europa no século IV a.C. Na América, o algodão já era utilizado por comunidades indígenas. A partir de 1860, a Inglaterra introduziu a variedade herbácea no Brasil, com o objetivo de incentivar a produção brasileira e diminuir a dependência relacionada aos Estados Unidos da América. A região escolhida foi São Paulo, onde se instituiu um mercado alternativo de algodão. O ciclo produtivo do algodão é ilustrado de forma genérica na Figura 2 (CORRÊA, 2003). A cotonicultura é composta por seis grandes etapas:

Plantio: abrangendo a seleção de variedades, preparo do solo e o plantio.

Colheita: etapas relacionadas com a colheita e armazenamento dos algodão na fase pré-beneficiamento.

Beneficiamento: etapas referentes ao descaroçamento e armazenamento dos subprodutos do algodão.

Controle de qualidade: realização de análises para qualificação e certificação da fibra do algodão;

Comercialização: etapa final envolvendo aspectos comerciais como venda e logística,

Figura 2 - Visão Geral dos processos na produção do algodão.

Fonte: (CORRÊA e COUTO, 2003).

2.3.1 Cotonicultura no Brasil

A expansão do plantio indica que o país também poderá assumir papel de destaque mundial na cultura do algodão. As plantações têm crescido especialmente em Estados como Mato Grosso, Bahia e Mato Grosso do Sul. Com alto grau de tecnologia, as lavouras de algodão apresentam resultados positivos em termos de produção e produtividade. Na Figura 3 é ilustrado um mapa com a distribuição das áreas de cultivo de algodão no Brasil, mostrando as regiões com maior concentração agrícola dessa cultura.

2.3.2 Análise de Solos

A quantidade de nutrientes exportados na cotonicultura é relativamente pequena, comparada a outras culturas de importância econômica. Devido a isso, a cotonicultura não é considerada um cultivo que degrade o local de plantio. Apesar disso, os problemas estão relacionados às medidas adotadas ao controle de pragas, normalmente feito por meio de queimadas. A verificação da qualidade do solo e eventuais correções são realizadas por intermédio de análises em talhões com tamanho máximo de 20 hectares com topografia homogênea, verificando características como cor e textura do solo. A quantidade de nutrientes disponíveis na área a ser cultivada é o fator que determinará uma adubação equilibrada; por isso, é importante conhecer previamente a quantidade de nutrientes extraídos e/ou exportados e quanto retornou para as áreas a serem cultivadas novamente, por meio dos restos culturais. Há, também, outros fatores que determinam a resposta das culturas à adubação, tais como: a dinâmica dos nutrientes no solo e o histórico de uso da área.

2.3.3 Seleção de Variedades

Richetti (2003) e Buainain (2007) relatam que o custo de insumos como sementes constitui uma das menores parcelas no plantio e, na cotonicultura, esse gasto corresponde, em média, de 2,3 a 3,0 % do custo total da lavoura. Sementes são manufaturadas por produtores e/ou empresas especializadas, cujo conteúdo são todos os genes que caracterizam a espécie a cultivar. A garantia da qualidade das sementes é estabelecida por entidades certificadoras que desempenham diferentes papéis no processo de produção dessas sementes, em cujo contexto são responsáveis pelo programa de melhoramento genético de determinada espécie. Nessa etapa da safra, o fundamental é estabelecer a qualidade e a variedade de semente mais adequada para o plantio nas áreas escolhidas.

2.3.4 Acompanhamento do Plantio

Segundo Richetti (2003), o algodão é uma planta oleaginosa e isso implica afirmar que, teoricamente, trata-se de uma cultura muito exigente quanto ao solo, preferindo áreas de cultivo de textura média, profundas, ricas em matéria orgânica,

permeáveis, bem drenadas e de boa fertilidade. Apesar desses requisitos, o algodão é uma cultura adaptável, no que se refere às condições de solo, podendo ser cultivado em diversas regiões com características físicas adversas e menos férteis, desde que sejam efetuadas as devidas correções. O manejo do solo constitui-se de práticas simples e indispensáveis para o bom desenvolvimento da cultura, compreendendo um conjunto de técnicas que, utilizadas racionalmente, proporcionam alta produtividade; entretanto, o seu mau uso pode levar à degradação e esgotamento, em curto prazo, podendo chegar, em casos extremos, à desertificação de áreas extensas.

Richetti (2003) ainda esclarece que o cerrado brasileiro, em particular, utiliza intensamente máquinas e implementos cada vez mais pesados e isso tem acelerado o processo de empobrecimento da matéria orgânica e a formação de encrostamento superficial e o surgimento de camadas adensadas na subsuperfície do solo. Esta, também chamada de camada compactada, responsável pela redução da taxa de infiltração, resulta no aumento da erosão, incidindo no incremento de perdas de nutrientes do solo. O uso de recursos tecnológicos no agronegócio proporcionou a modernização das lavouras, independente da cultura adotada e, no caso específico do algodão, é caracterizado pelas grandes áreas de plantio. A necessidade de produção em larga escala e de mecanização do cultivo favoreceu a utilização de equipamentos para a colheita por meio de colheitadeiras automotrizes. A produção mecanizada é extremamente vantajosa em relação à manual, pois os custos operacionais são reduzidos e há, também, melhoria na qualidade do produto colhido, maior rapidez no processo e baixo teor de impurezas.

3. ONTOLOGIAS

Esta pesquisa apresenta uma ontologia para formalizar o conhecimento utilizado na cotonicultura na região Centro-Oeste do Brasil, especificamente no Estado de Mato Grosso, dada a relevância no cultivo dessa cultura, sendo, atualmente, o maior produtor de algodão do país.

A pesquisa inicia-se com a investigação dos modelos formais para representação de conhecimento. O Quadro 1 expõe o resumo das características dos sete modelos analisados. O exemplar utilizado para a representação do conhecimento no segmento agrícola deve atender a algumas características básicas inerentes ao compartilhamento desse saber: hierarquia de dados, atributos, classes de dados, domínios de dados e representação gráfica. Nesse contexto, o modelo baseado em ontologia atende, de forma satisfatória, aos requisitos, mas sem detalhar a tecnologia ou especificações técnicas, restringindo-se à representação formal do conhecimento em um nível de abstração mais alto e favorecendo o intercâmbio da informação por meio das classes de dados mapeadas.

Quadro 1 - Características dos tipos de representação de conhecimento.

Recursos	Algoritmos	BPEL	BPMN	ER	Fluxograma	Ontologia	UML
Abstração Lógica	Sim	Sim	Sim	Não	Sim	Não	Sim
Atributos	Sim	Não	Não	Sim	Não	Sim	Sim
Cardinalidade	Não	Não	Não	Sim	Não	Não	Não
Chave Primária	Não	Não	Não	Sim	Não	Não	Não
Classes	Não	Sim	Não	Não	Não	Sim	Sim
Domínio	Sim	Não	Não	Sim	Não	Sim	Sim
Eventos	Não	Sim	Sim	Não	Sim	Não	Sim
Gráficos	Não	Sim	Sim	Sim	Sim	Sim	Sim
Herança	Não	Não	Não	Não	Não	Não	Sim
Hierarquia	Não	Não	Não	Não	Não	Sim	Não
Instanciamento	Não	Não	Não	Não	Não	Não	Sim
Processos	Não	Sim	Sim	Sim	Sim	Não	Não
Pseudolinguagem	Sim	Não	Não	Não	Não	Não	Não
Reusabilidade	Não	Sim	Sim	Não	Não	Não	Sim

As ontologias oferecem um meio de compartilhamento e reuso do conhecimento a respeito de um domínio específico, sendo esse modelo formal empregado em diversas áreas, tais como aplicações com *web* semântica, *e-commerce*, medicina, indústria automotiva, gestão financeira ou outra área relacionada à gestão de conhecimento (LI, 2005). Um exemplo genérico a respeito de ontologias é apresentado na Figura 4 por meio da Árvore de Porfírio utilizada

para representar o conhecimento sobre substâncias. Ontologias também definem os termos utilizados para descrever e representar uma área de conhecimento que pode ser compartilhado por pessoas, gerenciadores de banco de dados e aplicações específicas (CRUZ, 2004). No contexto desta pesquisa, a ontologia será usada como ferramenta para a representação das classes de dados na cotonicultura.

Figura 4 - Árvore de Porfírio.

Fonte: (COHEN, 2007).

3.1 Classificação

De acordo com Cruz (2004), ontologias também podem ser utilizadas para descrever, de forma hierárquica, estruturas de dados. A partir de uma ontologia, é possível criar um padrão para fornecer metadados para a descrição e intercâmbio de dados entre os processos de um domínio específico. A identificação das classes de dados chave para os processos de produção é discutida na pesquisa conduzida por Fileto (2011). A Figura 5 apresenta uma classificação geral sobre os tipos de ontologias de acordo com escopos de uso, conforme Guarino (1998).

Figura 5 - Tipos de ontologias.

Fonte: Guarino (1998).

Ontologias de alto nível: representam conceitos genéricos que independem de um domínio particular, como: espaço, evento e tempo. De acordo com Guarino (1998), a representação ontológica de alto nível não aborda, de forma detalhada, conforme ilustrado na Figura 6, os elementos de um cenário específico, pois o objetivo é representar, conceitualmente, os elementos mais relevantes para o intercâmbio de conhecimentos simples.

Figura 6 - Fragmento de ontologia de alto nível para comercialização agrícola.

Ontologias de domínio: para Musen (1995), ontologias de domínio são descrições formais de classes de conceitos e as relações entre esses conceitos que descrevem uma área de aplicação. A partir dessa definição, na Figura 7 é ilustrada um fragmento da hierarquia de processos para o mapeamento da informação no beneficiamento agrícola, tanto de grãos quanto do cotonicultor. Nesse tipo de ontologia, são apresentados conceitos sobre um domínio específico por meio de um vocabulário, partindo de um domínio mais genérico, como: agricultura, computação, medicina, entre outros, para um domínio mais específico.

Figura 7 - Fragmento de ontologia de domínio cultivo e beneficiamento agrícola.

Ainda de acordo com Zong-yong et al. (2007), ontologias baseadas em domínios desempenham um papel importante na engenharia de requisitos, contribuindo de forma eficiente para a formalização de conhecimento, possibilitando, também, o uso desse conhecimento de forma eficaz. Por meio dos modelos de conhecimento de domínio, usuários podem sugerir melhorias de acordo com necessidades mais específicas, contribuindo, de forma mais precisa, com a compreensão de requisitos por analistas.

Ontologias de tarefas: para Ikeda et al. (1999), ontologias baseadas em tarefas são como um sistema de semântica de recursos para representar o significado da descrição de resolução de problemas. A principal vantagem do seu uso reflete em fornecer formas primitivas em vocabulários, com termos cativantes, para que usuários possam facilmente descrever a sua própria resolução de problemas e processos. Isso possibilita a representação de conceitos para a solução de obstáculos, independente do domínio no qual ocorrem, ou seja, definir vocabulários de acordo com as atividades ou tarefas genéricas, como a representação do conhecimento em um processo industrial ou representação de uma operação mecanizada, conforme ilustrado na Figura 8.

Figura 8 - Fragmento de ontologia de tarefas para gerenciamento administrativo.

Ontologias de aplicação: de acordo com Malone (2010), ontologias de aplicação são projetadas para um uso específico ou foco de aplicação cujo escopo é especificado por meio de casos de emprego com possibilidade de realização de testes. A ontologia de aplicação, em geral, usa ou faz referência a regras para a construção de classes e as relações entre as classes ontológicas. São utilizadas na modelagem de experiências entre domínios, para anotações ou visualizações de dados para grupos específicos de usuários. Neste caso, os conceitos, geralmente, descrevem funções realizadas por entidades de um domínio responsável por uma definida atividade, como: ontologia para aplicação em manejo no cultivo de determinada cultura. Na Figura 9 é ilustrado um fragmento de ontologia para a interoperabilidade de dados entre domínios específicos.

Figura 9 - Fragmento de ontologia de aplicação para interoperabilidade de dados.

3.2 Classes e Atributos

Classes e atributos são características presentes em grande parte dos modelos para formalização de cenários e modelagem de dados. A possibilidade de identificar entidades, singularmente com as respectivas características, favorece a qualidade e a precisão do modelo representativo de um cenário do mundo real, conforme ilustrado na Figura 10. O modelo utilizado na pesquisa fornece recursos para a representação de características como: fazenda, quadra, blocos, sub-blocos, com a possibilidade de agregar atributos como: descrições, áreas, localizações entre outros.

Figura 10 - Exemplo de classes e atributos.

3.3 Taxonomia

A comunicação entre processos geradores de dados a fonte da troca de informações a respeito de entidades e relações entre um emissor e um receptor. Tais informações são formuladas em alguma linguagem que consista de símbolos arranjados de um modo definido e que não sejam significativos em si. O significado de uma simbologia precisa ser explicitado, especificando a interpretação e a que se destina, ou seja, mencionando a qual entidade ou relação se refere. Nesse contexto, metadados são indispensáveis para a construção do conhecimento formal sob uma óptica simbólica e intercambiável. A seguir, são apresentadas algumas aplicações para ontologias em cenários distintos.

Aplicações lógicas: segundo Bittner (2004), aplicações lógicas referem-se à representação de teorias lógicas, os termos da terminologia, cuja semântica deve ser especificada, pois aparecem como nomes, predicado e símbolos de relação da linguagem formal. Axiomas e definições lógicas são adicionados para expressar associações entre as entidades, classes e relações concebidas por esses símbolos. O Quadro 2 é ilustrado a construção de um fragmento de ontologia por meio de relações entre fatos no contexto agrícola.

Quadro 2 - Fragmento de ontologia baseado em lógica.

```

predicates
  fazenda(symbol)
  quadras(symbol,symbol)
  blocos(symbol,symbol,integer)
  ontologia(symbol)

clauses
  fazenda(mtgrosso) .
  quadras(mtgrosso,th01) .      quadras(mtgrosso,th02) .
  quadras(mtgrosso,th03) .      quadras(mtgrosso,th04) .
  blocos(mtgrosso,th01,12) .    blocos(mtgrosso,th01,10) .
  blocos(mtgrosso,th01,11) .    blocos(mtgrosso,th02,10) .
  blocos(mtgrosso,th02,12) .    blocos(mtgrosso,th02,11) .

```

$ontologia(N) :- fazenda(N), quadras(N, _), blocos(N, _, _).$

Aplicações não lógicas: ontologias não lógicas não especificam semântica sistêmica, restringindo as interpretações possíveis dos termos por meio de axiomas lógicos. Nesse contexto, é especificado o significado de uma terminologia por meio da interpretação dos vocábulos individualmente. A redução da ambiguidade dos mesmos é obtida pelo uso de símbolos com significados únicos. Na Figura 11 é ilustrada uma representação de blocos na produção de algodão bruto.

Figura 11 - Representação de ontologia não lógica.

Interoperabilidade: são ontologias utilizadas para o compartilhamento de aplicações e na comunicação entre processos. A semântica dessa terminologia é indicada por meio de um metapadrão e todas as aplicações envolvidas podem se comunicar, de forma inequívoca, empregando a mesma nomenclatura. Em aplicação interna, pode-se usar uma terminologia diferente da norma estabelecida e, nesse caso, mapeamentos de transformação devem ser definidos para proporcionar a comunicação de forma eficiente (BITTNER, 2004). Na Figura 12 é ilustrada a formalização de uma ontologia aplicada à interoperabilidade na agricultura. A ilustração mostra um fragmento de ontologia, representando fazenda, quadras e blocos com os respectivos atributos.

Figura 12 - Fragmento de ontologia aplicado à cotonicultura.

3.4 Intercâmbio de Dados

Durante vários anos, operações comerciais eram realizadas por meio de movimentações financeiras baseadas em documentos impressos, contendo informações necessárias para transmissão de dados a fim de efetuar transações comerciais. Posteriormente, os documentos começaram a assumir padrões, como notas fiscais, notas de créditos e ordens de serviços. Inicialmente, os apontamentos não foram planejados para seguir qualquer *layout* ou padrão, por não representar algo essencial, pois os destinatários sempre eram pessoas com capacidade de ler, interpretar e racionalizar.

A partir da década de 1950, os computadores passaram a ser usados por grandes empresas em setores como contabilidade e folha de pagamento. Ao longo das décadas seguintes, essa nova tecnologia assumiu papel estratégico em outros setores e funções envolvidas na produção, administração e todas as outras áreas comerciais. Apesar dessa tendência crescente, havia uma grande demanda por documentos impressos em vários formatos, sem *layouts* consolidados. Esse cenário apresentava uma necessidade iminente de definir regras para emissores e receptores de documentos.

Computadores não interpretam facilmente documentos impressos e criar recursos para esse tipo de tarefa é quase impossível. Devido a isso, empresas receptoras contratavam pessoal para desempenhar a função de intérpretes das informações contidas nos documentos recebidos, mesmo em sistemas informatizados (DATA INTERCHANGE, 2005).

Nesse contexto, o uso de tecnologia da informação para apoiar o intercâmbio de dados, dentro e entre as organizações, tem aumentado de forma significativa. Exemplo disso é a troca de conhecimento estruturada entre as empresas,

geralmente referida como *Electronic Data Interchange* ou apenas EDI (O'CALLAGHAN, 2005).

A proposta do uso de *EDI* é colaborar com a redução de problemas relativos à comunicação entre processos; o intercâmbio de dados em sistemas gestores de Recursos Humanos, *ERPs* e sistemas verticais especializados ilustram esse e os problemas supracitados relativos à dificuldade de permuta de dados. Em geral, soluções computacionais para *ERP* não incorporam gestão de RH e a razão para essa verticalização está no *core business* de cada empresa. Com isso, surge um *gap* entre dois sistemas que deveriam estar compartilhando dados comuns e o resultado é redundância de elementos e retrabalho para setores da mesma empresa.

Outro ponto essencial está na necessidade de dados homogêneos para viabilizar o uso de ferramentas gestoras de informações. Soluções como *Business intelligence (BI)*, *Knowledge Discovery in Database (KDD)* e *Data Warehouse*, entre outros, demandam dados com qualidade para a criação de informações úteis. Apesar do pouco uso de soluções mais complexas do ponto de vista de gestão de dados, há a necessidade real de homogeneizá-los, mesmo em aplicações com menor grau de complexidade, como *ERPs*. O Gráfico 1 apresenta uma visão das soluções computacionais utilizadas na agricultura no Centro-Oeste brasileiro e foram obtidas junto a gestores de tecnologia em 12 empresas relacionadas com o agronegócio. Os dados mostram pouca adoção de sistemas gestores de informação mais sofisticados, sendo que a gestão corporativa de dados no segmento agricultura é feita, em sua maioria, por sistemas com foco operacional, limitando-se à inserção e manutenção de dados.

Gráfico 1 - Soluções corporativas utilizadas no agronegócio.

4. METADADOS

Um passo relevante para a padronização e interoperabilidade de dados no agronegócio está na criação de metadados, sendo que algumas vantagens são:

- Padronização da descrição de conjuntos de dados,
- Qualidade dos conjuntos de dados representados e
- Interoperabilidade entre processos e dispositivos físicos distintos.

Um ponto considerável explorado na pesquisa é o uso de metadados no agronegócio, especificamente, na produção do algodão. A proposta é criar um padrão para o transporte de dados nas principais etapas referentes ao plantio, colheita e beneficiamento. Conceitualmente, metadados podem ser definidos como dados a respeito de dados e na Figura 13 é ilustrada uma visão estrutural geral a respeito da composição de um metadado. O item de um metadado determina o significado de um dado específico, usualmente, representando uma informação inteligível por pessoas e sistemas computacionais. Segundo a *National Information Standards Organization-NISO* (2004), os metadados facilitam o entendimento dos relacionamentos e o uso das informações dos dados.

Figura 13 - Visão geral de metadados.

Fonte: Niso (2004)

4.1 Padrões de Metadados na Agricultura

Segundo Coughlin (2010), usar padrões para acompanhar e descrever a essência de mídia de dados melhora o intercâmbio de dados entre processos e alguns segmentos mais específicos fazem uso de padrões próprios com esse propósito, como *agroxml* e *agxml*.

O uso de metadados para aplicações agrícolas é realizado, utilizando os mesmos critérios e estruturas adotados em outras áreas. Todavia, padrões genéricos podem dificultar o intercâmbio e a recuperação de dados, pois cada fabricante define estruturas proprietárias para a troca de mensagens. Apesar de haver uma padronização na linguagem empregada, usualmente o *XML*, não há conformidade nas estruturas de dados para tratamentos posteriores, como atualizações e recuperações de dados. Problemas como esses serviram de iniciativa para o surgimento de algumas propostas que definissem novos padrões adequados às necessidades e particularidades do agronegócio (VALKENGOED, 2007).

4.1.1 Dublin Core

O *Dublin Core* é um vocabulário de quinze propriedades para uso na descrição de recursos. O nome *Dublin* é consequência da origem em um *workshop* em 1995, ocorrido em *Dublin, Ohio*; e *Core* representa que os elementos são amplos e genéricos, utilizáveis para descrever uma vasta gama de recursos. Os elementos do *Dublin Core* fazem parte de um conjunto maior de vocabulários de metadados e as especificações técnicas mantidas pelo *Dublin Core Metadata Initiative - DCMI* (INITIATIVE D.C., 2010).

Desde 1998, algumas recomendações na *web* semântica evoluíram para incluir a atribuição de domínios formais e faixas, além de definições em linguagem natural. Domínios especificam os tipos de recursos e valores podem ser associados a uma determinada propriedade. Quando uma exata propriedade é encontrada, um pedido de inferência pode utilizar a informação nos domínios e faixas atribuídas a uma propriedade, a fim de fazer inferências sobre os recursos assim descritos. Desenvolvedores podem escolher livremente o uso dessas particularidades em função das necessidades do aplicativo (COMMUNITIES, 2010).

Característica(s) Favorável(is): o Dublin Core é um dos padrões de metadados mais aceitos no meio acadêmico para formalização de dados relacionados com biodiversidade.

Característica(s) Desfavorável(is): o padrão Dublin Core foi idealizado exclusivamente para o tratamento de dados ligados à biodiversidade; isso o torna específico quanto ao escopo de uso para outros segmentos.

4.1.2 Darwin Core

O nome *Darwin Core* foi proposto por Allen Allison em 1998 com semelhança conceitual ao *Dublin Core*. O *Darwin Core* foi posteriormente expresso em um documento *Zbig*. O padrão é estruturado a partir de um núcleo, representando o corpo de normas, que inclui um glossário de termos, semelhante a propriedades, elementos, colunas ou atributos, destinado a facilitar o compartilhamento de informações sobre a diversidade biológica, fornecendo definições de referência, exemplos e comentários. O *Darwin Core* se baseia principalmente nas taxas e ocorrências na natureza de determinadas espécies, documentadas por meio de observações, famílias, amostras e informações relacionadas (TDWG, 2007).

Característica(s) Favorável(is): a proposta do *Darwin Core* é semelhante ao *Dublin Core*, sendo uma evolução deste, com propósito de atender à formatação de dados ligados à biodiversidade, de forma mais específica que o *Dublin Core*.

Característica(s) Desfavorável(is): as semelhanças quanto ao *Dublin Core* também associam-se a desvantagens. A principal delas é a verticalização do padrão, tornando o uso muito específico.

4.1.3 AgMES

A iniciativa *Agricultural Information Management Standards*, ou *AgMES*, tem o objetivo de contemplar questões de semântica e normas no domínio da agricultura, relacionadas com descrição, descoberta de recursos, interoperabilidade e intercâmbio de metadados para os diferentes tipos de informação. Segundo a *Agricultural Information Management Standards*, o grupo atua como um guarda-chuva sob o qual novos elementos ou refinamentos podem ser usados para

diferentes recursos em todas as áreas tecnológicas, relevantes para produção de alimentos, nutrição e desenvolvimento rural (WANG et al., 2008).

Característica Favorável: o objetivo é prover normas semânticas no segmento agrícola para a melhoria do intercâmbio de metadados que favoreçam a criação de soluções tecnológicas.

Característica Desfavorável: a iniciativa *AgMES* tem características semelhantes aos demais padrões verticais para agricultura, com uso restrito ao agronegócio.

4.1.4 AGRIS

O *AGRIS* é um padrão criado, especificamente, para melhorar a descrição, o intercâmbio e posterior recuperação de dados referentes à produção agrícola. É um modelo que permite o compartilhamento de informações entre diferentes sistemas bibliográficos e baseia-se em conhecimentos aceitos e padrões de metadados. As orientações também fornecem as melhores práticas recomendadas para a catalogação e indexação de assunto. O *AGRIS* é um passo importante para o formato de alta qualidade em metadados de média complexidade em um aplicativo de troca de dados independente (NATLACEN et al.; BARTOL, 2010).

Característica Favorável: uma característica importante, presente no *AGRIS*, está na própria estrutura para descrever recursos, não se limitando a dados isolados por meio de *tags*, e é acessível a objetos em diferentes formatos, tanto em textos simples quanto em bases de dados.

Característica Desfavorável: o *AGRIS* não é extensível a outras áreas, restringindo-se a um segmento específico.

4.1.5 Agrovoc

O dicionário de sinônimos *AGROVOC* é um vocabulário multilíngue estruturado e controlado, criado para armazenar terminologias de todas as áreas temáticas no domínio da agricultura, silvicultura, pesca, alimentos e domínios conexos, como meio ambiente. O *AGROVOC* foi instituído, inicialmente, como um dicionário de sinônimos, mas agora está evoluindo para um serviço de ontologia. Esse enriquecimento está ocorrendo em várias iniciativas na *web* semântica e pode ser definido como um sistema semântico que contém termos, definições de termos e

especificação das relações entre termos. O sistema semântico pode ser referido como "*Service Ontology*", criado e mantido pela *Food and Agriculture Organization* ou *FAO*, disponível desde o início dos anos de 1980, com atualizações constantes (SANCHEZ-ALONSO et al., 2009).

Característica Favorável: o uso de uma fonte única para consulta de vocábulos facilita a comunicação entre processos. O emprego de um padrão de termos possibilita, até mesmo, a criação de serviços para usuários finais por meio de *webservices*.

Característica Desfavorável: trata-se de um dicionário de termos em um sistema gerenciador de banco de dados, o que limita o uso nas aplicações no agronegócio, de uma forma geral.

4.1.6 AGXML

O AGXML é um padrão desenvolvido para atender à formalização de dados no segmento agrícola de grãos, favorecendo, também, o processamento de informações empresariais e entidades relacionadas (AGXML, 2011). O grupo de trabalho empenha-se no desenvolvimento de padrões para a troca eficiente de informações eletrônicas no agronegócio e tem como metas:

- Desenvolver normas que possam ser usadas para melhorar a eficiência dos processos internos e externos, incluindo o envio de documentos por meio da *Internet*.
- Definir documentos de negócios envolvidos no volume da produção de grãos e oleaginosas, abrangendo o conteúdo de documentos, esquemas XML e um glossário.
- Alinhar-se a outros órgãos técnicos relacionados com agronegócio e padrões.
- Ser organizacionalmente auto-suficiente, maior do que qualquer empresa.
- Tornar-se um padrão que seja versátil e extensível (AGXML, 2011).

Características Favoráveis: a estrutura do *AgXML* proporciona o mapeamento com precisão dos dados referentes à produção da cultura da soja. O vocabulário é detalhado e abrangente a todos os processos desse segmento agrícola. Outro ponto importante é o fato de utilizar a linguagem XML como base, o que proporciona transparência e facilidade no uso.

Característica Desfavorável: o *agXML* é um padrão vertical, dificultando ou inviabilizando o uso para representação de processos em outras culturas.

4.1.7 AGROXML

O *AGROXML*, criado em 16 de setembro de 2003 como um conceito e marca e, em 14 de abril de 2004, foi apresentado publicamente, pela primeira vez, no final do simpósio "Documentação Pública e Rastreabilidade na Agricultura" na *Fachhochschule Bingen University of Applied Sciences*. Em 20 de julho de 2004, foi realizada, na Universidade de *Hohenheim*, uma reunião de consulta. Inicialmente, com um grupo reduzido de representantes de empresas de *software* agrícola, cuja proposta era apresentar e discutir o *AGROXML*. Em 30 de setembro de 2004, realizou-se a primeira reunião do grupo de trabalho em *AGROXML*, mostrando o retorno positivo obtido com sua aplicação (SCHMITZ et al., 2009).

De acordo com Martini et al. (2009), o *AGROXML* fornece um método de armazenamento de dados estruturados por assuntos agrícolas. Os dados podem ser armazenados ou trocados entre os diferentes participantes e inclui elementos particulares da cultura. O *agroXML* é composto por doze módulos que disponibilizam *tags* necessárias para atender, de forma genérica, às especificidades do agronegócio.

Característica Favorável: a proposta do *agroXML* é a criação de um padrão comum de interoperabilidade para troca de dados entre dispositivos e processos no agronegócio. Apesar de o padrão ter sido desenvolvido de forma quase independente, está presente em diversas discussões no segmento agrícola.

Característica Desfavorável: o *agroXML*, mesmo possuindo doze módulos abrangendo os principais grupos de dados, não atende a todos os requisitos na representação de dados em todos os segmentos do agronegócio. Essa limitação deve-se à existência de particularidades em diferentes regiões para a mesma cultura.

4.2 Padrão para Cotonicultura

Conforme descrito no item 1.8.1, existem diversos padrões de metadados para aplicações na agricultura; entretanto, não há um padrão específico para atender às particularidades da cotonicultura, como: controle de qualidade de sementes, gestão de recursos hídricos, acompanhamento do plantio e, principalmente, a etapa de beneficiamento. Esta pesquisa propõe um padrão de metadados, contemplando as características da cotonicultura. A formalização do padrão será realizada, utilizando a linguagem *XML*. A escolha da linguagem deve-se à sua extensibilidade na criação de estruturas para armazenamento de dados de forma totalmente flexível quanto à hierarquia de dados. A linguagem *XML* também apresenta viabilidade técnica e extensibilidade para ferramentas de gerenciamento de dados e formalização de ontologias. Assim, no Gráfico 2 são ilustradas as principais linguagens para formalização de ontologias de acordo com Denny (2004). A sua estrutura permite legibilidade e transparência no uso das *tags* disponibilizadas para criação de soluções, destinadas à homogeneização de dados.

Gráfico 2 - Linguagens para formalização de ontologias.

4.3 Dimensões de Qualidade de Metadados

Metadados são criados para atender deficiências relativas à padronização de dados em quaisquer segmentos, como indústria e agricultura. O uso de metadados de forma automática ou manual por meio de gestores de informações não familiarizados com catalogação, indexação, ou controle de vocabulário, resulta em problemas com a qualidade dos dados. Ocasionalmente, a ausência ou inconsistência de elementos obrigatórios, a sintaxe de esquemas pode conter erros impeditivos para que os metadados sejam processados corretamente. O conteúdo da terminologia de metadados também pode ser instável, tornando difícil localizar informações relevantes. A *National Information Standards Organization* define seis recomendações para verificação da qualidade de padrões de metadados. O propósito não é definir regras, mas compartilhar o conhecimento obtido em experiências adquiridas por meio das características de outros padrões de metadados, em diversas áreas como ciências humanas e agricultura:

- Coleções de Dados: metadados devem suportar coleções de dados com domínio compatível com a necessidade de armazenamento.
- Interoperabilidade: resulta na possibilidade do padrão em atender plataformas distintas de forma transparente.
- Vocabulários Controlados: o padrão deve definir o que, onde, quando e a quem se destina; padrões excessivamente abertos são propícios à desconstrução da estrutura original dos mesmos.
- Condições e Termos de Uso: é importante que haja boa documentação a respeito de possibilidades e restrições técnicas e legais quanto ao uso do padrão.
- Arquivamento: armazenamento, persistência, identificação única, entre outras, devem ser características obrigatórias e de fácil compreensão.
- Gestão Contínua: manutenção e evolução periódica do padrão por grupos de trabalho.

No Quadro 3 é ilustrado um comparativo, verificando os padrões de metadados abordados na pesquisa relacionada às características expostas pela *National Information Standards Organization*:

Quadro 3 - Análise de características dos padrões de metadados.

Métricas	<i>Dublin</i>	<i>Darwin</i>	<i>AgMES</i>	<i>AGRIS</i>	<i>Agrovoc</i>	<i>agXML</i>	<i>agroXML</i>
Coleções de Dados	Sim	Sim	Sim	Sim	Não	Sim	Sim
Interoperabilidade	Sim	Sim	Sim	Sim	Não	Sim	Sim
Vocabulários Controlados	Sim	Sim	Sim	Sim	Não	Sim	Sim
Condições e Termos de Uso	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Arquivamento	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Gestão Contínua	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Estruturalmente, a maioria dos padrões analisados atende aos requisitos sugeridos pela *National Information Standards Organization (NISO)*. Entretanto, o *Agrovoc* possui uma estrutura disposta em tabelas de dados, diferente dos demais padrões formalizados em linguagem de marcação e, devido a isso, não considera alguns itens.

4.4 Contribuições Multidisciplinares

Conforme contextualizado anteriormente, existem poucos padrões para o intercâmbio de dados na cotonicultura e os padrões existentes não proporcionam a estrutura adequada para atender aos requisitos referentes a dados gerados na cotonicultura. Por essa razão, é proposto um padrão de metadados utilizando o conceito ontológico, por meio de uma ontologia de domínio em razão da possibilidade de representação de classes de forma hierárquica. Nesse sentido, foi realizada uma pesquisa com características multidisciplinares, envolvendo as áreas de agrárias, computação e engenharia; dessa forma, as contribuições observadas atendem às áreas supracitadas com as seguintes particularidades:

Ontologia: a construção da ontologia proporcionará o desenvolvimento de modelos, tanto para armazenar quanto para apresentar dados em soluções com

foco na gestão de processos, permitindo, também, o reuso da metodologia aplicada em outros cenários como pecuária e agroindústria.

Qualidade de dados: o uso da hierarquia de dados, proposta pelo padrão de metadados, possibilita melhora direta na qualidade de algumas dimensões de dados, com contribuição direta a aspectos relacionados à engenharia de *software*, subsidiando, também, aspectos que envolvam o desenvolvimento de aplicações com diversos paradigmas de programação.

Homogeneização de dados no contexto computacional: o aporte relacionado à homogeneização reflete em resultados interdisciplinares na geração de informações para sistemas de tomada de decisão (*BI*, *data mining*, entre outros), com foco intermediário em recursos computacionais e proporciona ações para implementações em engenharia da computação como subsídios para construção de sistemas para classificações diversas.

Metadados: no contexto da pesquisa, padrões de metadados são utilizados para a formalização da ontologia por meio de regras e hierarquias necessárias à definição de infraestruturas usadas, tanto para criação de interfaces entre sistemas legados quanto para intercâmbio de dados entre aplicações com estruturas heterogêneas na gestão agrícola.

Homogeneização de dados no contexto agrícola: o mapeamento dos fluxos de dados entre os processos possibilita a visão bidirecional do comportamento de todas as entradas e saídas; dessa forma, a recuperação de dados e a geração de informações como: agrupamentos de ordens de colheita, estoque em almoxarifado, previsões de recursos humanos e projeções de compras, tornam-se mais dinâmicas, eliminando a necessidade da criação de aplicações específicas para cada situação nesse cenário. Na Figura 14 é ilustrado o fluxo de desenvolvimento da pesquisa, partindo da obtenção dos requisitos e, posteriormente, o mapeamento da ontologia formalizada no padrão de metadados proposto, utilizando uma prova de conceito para validação.

Figura 14 - Visão geral do fluxo de desenvolvimento da pesquisa.

A contextualização do cenário da pesquisa aborda, conceitualmente, ontologias e metadados como recursos para contribuir com a melhoria da integração de dados e consequente qualidade dos mesmos. Os conceitos abordados aplicam-se, nesse contexto, a um segmento agrícola específico com possibilidade de utilização da metodologia para mapeamento dos processos, construção da ontologia e criação do padrão de metadados em outras áreas, como pecuária e logística, sendo essas áreas transversais ao escopo abordado.

5. MATERIAIS E MÉTODOS

A pesquisa inicia-se com a investigação dos modelos formais para representação de conhecimento. O Quadro 4 expõe o resumo das características dos oito modelos para a representação analisada. O modelo formal utilizado para essa representação no segmento agrícola deve atender a algumas características básicas inerentes ao compartilhamento desse conhecimento: hierarquia de dados, atributos, classes de dados, domínios de dados e representação gráfica. As características supracitadas referem-se a requisitos necessários para a representação do modelo baseado em hierarquia de classes de dados. Nesse contexto, o modelo baseado em ontologia atende, de forma satisfatória, aos requisitos, mas sem detalhar a tecnologia ou especificações técnicas, restringindo-se à representação formal do conhecimento em um nível de abstração mais alto e favorecendo o intercâmbio da informação por meio das classes de dados mapeadas.

Quadro 4 - Características dos tipos de representação de conhecimento.

Recursos	Algoritmos	BPEL	BPMN	ER	Fluxograma	Ontologia	UML
Abstração Lógica	Sim	Sim	Sim	Não	Sim	Não	Sim
Atributos	Sim	Não	Não	Sim	Não	Sim	Sim
Cardinalidade	Não	Não	Não	Sim	Não	Não	Não
Chave Primária	Não	Não	Não	Sim	Não	Não	Não
Classes	Não	Sim	Não	Não	Não	Sim	Sim
Domínio	Sim	Não	Não	Sim	Não	Sim	Sim
Eventos	Não	Sim	Sim	Não	Sim	Não	Sim
Gráficos	Não	Sim	Sim	Sim	Sim	Sim	Sim
Herança	Não	Não	Não	Não	Não	Não	Sim
Hierarquia	Não	Não	Não	Não	Não	Sim	Não
Instanciamento	Não	Não	Não	Não	Não	Não	Sim
Processos	Não	Sim	Sim	Sim	Sim	Não	Não
Pseudolinguagem	Sim	Não	Não	Não	Não	Não	Não
Reusabilidade	Não	Sim	Sim	Não	Não	Não	Sim

A proposta da pesquisa é criar uma ontologia baseada em domínio para representação do conhecimento relacionado com um escopo agrícola específico e previamente estabelecido, envolvendo processos geradores de dados heterogêneos na cotonicultura, de acordo com as pesquisas apresentadas por Sofiatti (2009) e Bortolini (2005).

A identificação das classes segue uma sequência hierárquica, organizada de forma cronológica, contemplando uma abordagem que obedece à organização gerencial do ponto de vista da produtividade, de acordo com as características estruturais referentes à ontologia de domínio, adequada ao cenário da pesquisa, conforme discutido no item 1.6.1. Dessa forma, o ponto de partida para construção da ontologia para cotonicultura é a área de cultivo, conforme sugerido por Richetti (2003) e Buainain (2007) em trabalhos compilados para a Empresa Brasileira de Pesquisa Agropecuária e Ministério da Agricultura, Pecuária e Abastecimento. Com isso, todo o planejamento processual posterior, referente à seleção da variedade mais adequada para composição do solo, correções de solo, aplicação de insumo, aplicação de defensivos, entre outros, está diretamente vinculado às características do espaço geográfico de origem. A delimitação da área ou fazenda é a base e a área de trabalho para execução das atividades posteriores será, também, a origem para os produtos e subprodutos dessa atividade agrícola, sob o ponto de vista ontológico, conforme Yamaguchi (1999). Nesse contexto, a fazenda será definida como a classe inicial ou ponto de partida para o mapeamento das demais classes de dados hierarquicamente inferiores, cuja metodologia contempla a identificação das classes intermediárias, desde o plantio até o elemento final desejado para a representação do conhecimento, no cenário representado pelo beneficiamento e comercialização da fibra. É importante destacar que os princípios básicos da administração, aplicados à indústria e comércio, são válidos para a agricultura, segundo Reichert (2001). Entretanto, deve-se ressaltar que algumas características a diferenciam dos demais segmentos. A classe de dados Fazenda é descrita no modelo como uma raiz para as demais classes. Posteriormente, a área delimitada para cultivo será subdividida em áreas menores, de acordo com a necessidade de cuidados diferenciados. Conforme sugerido por Ferguson (2009), a solução para o problema está na criação de áreas de manejo diferenciadas, sendo essas subdivisões a segunda classe hierarquicamente inferior à fazenda, também denominada de quadra ou talhão. Análoga à justificativa dos dois primeiros elementos identificados para construção da ontologia, segue a mesma metodologia para o mapeamento das demais classes.

A investigação a respeito da homogeneização de dados na cotonicultura teve início com a análise de características de sistemas computacionais utilizados para gestão e rotina administrativa. O diagnóstico mostrou as deficiências na geração dos dados, e, por consequência, as falhas no intercâmbio dos mesmos. Nesse contexto, as insuficiências são resultado do isolamento entre sistemas computacionais em setores distintos de uma mesma empresa, devido à fraca integração de dados, ocasionando o empobrecimento dos mesmos.

O conjunto de atributos necessários para a construção da ontologia foi obtido por meio de estudos realizados sobre conjuntos de processos em várias etapas da cotonicultura. Posteriormente, obteve-se como resultado a descrição dos termos mais frequentes e, além das características identificadas, também foram verificadas possíveis hierarquias entre os processos transformados em conceitos. Todas as informações detectadas durante os estudos do cenário foram mapeadas manualmente em uma ontologia de domínio com a descrição dessa estrutura, bem como as relações importantes entre as mesmas. A ontologia de domínio foi utilizada, visando identificar dois tipos específicos: conceitos sobre a estrutura de classes derivada dos processos da cotonicultura e o domínio nesse escopo. Os conceitos referentes à estrutura possibilitam a identificação das características mais relevantes, como as classes de dados que, nesse contexto, definem a sequência processual da cotonicultura. No caso da pesquisa realizada, os processos foram analisados de acordo com referências bibliográficas e fontes complementares, obtidas em visitas técnicas e entrevistas acerca da composição de todos os envolvidos na cotonicultura. Por exemplo, com as seguintes características: áreas de cultivo, irrigação, beneficiamento, entre outros. De forma semelhante, os conceitos de domínio são empregados na identificação de características do conteúdo ou dados gerados com o foco utilizado no cenário inicial e, neste contexto, a cotonicultura.

A metodologia adotada para a identificação das classes de dados na ontologia proposta parte do reconhecimento dos processos e da conversão dos mesmos em classes de dados correspondentes. Essas classes não representam ações desempenhadas pelos processos e o seu uso está direcionado aos dados

gerados em cada processo, tornando explícitas, também, as relações hierárquicas com outras classes subordinadas, conforme ilustrado na Figura 12. É importante destacar que soluções computacionais agrícolas, usualmente, não são projetadas para o compartilhamento de dados com aplicações externas ao domínio de aplicação. Para ilustrar isso, nos Gráficos 3 e 4 são ilustrados os resultados da coleta de dados referentes à capacidade de sistema em exportar, importar e disponibilizar recursos para conexão externa com aplicações. Essa coleta foi realizada tanto no Brasil quanto em empresas europeias, asiáticas e americanas, com investigação em quatorze empresas brasileiras e quatro empresas nos demais países, a coleta de dados foi realizada diretamente ao pessoal técnico de cada empresa por meio de questionário eletrônico. O resultado dos dados relativos ao Brasil apresenta pouca acessibilidade de interação entre sistemas. Em geral, os recursos relacionados com intercâmbio de dados limitam-se à exportação e importação em formatos comuns. Outro ponto importante é a tecnologia utilizada para a disponibilização e integração com outros sistemas, habitualmente realizada por meio manual, com operações de exportações em sistemas de origem e importação em sistemas de destino. A ausência de regras específicas para a representação dos dados na agricultura dificulta o desenvolvimento de soluções tecnológicas que atendam às demandas desse segmento.

Gráfico 3 - Empresas brasileiras

Gráfico 4. Empresas estrangeiras

Um ponto primordial para a pesquisa está na delimitação das classes com maior importância. Para isso, foi utilizado como referência inicial o trabalho desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária, EMBRAPA e compilado em documento por Richetti et al (2003). Também foi aproveitado o estudo supracitado para estabelecer, de forma consistente, o grau de granularidade necessário à representação dos dados para esse segmento. Outras fontes de dados foram usadas complementarmente, sendo, então, categorizadas em: acadêmica,

certificadora, comercial, governamental e de tecnologia, conforme ilustrados no Quadros 5.

Quadro 5 - Locais de visitas técnicas - Acadêmicas.

Instituição	Local	Motivo/Contribuição
Economia UFMT	Cuiabá Mato Grosso	Coleta de dados referente a aspectos sociais da cotonicultura.
Engenharia Agrícola/UFMT	Rondonópolis Mato Grosso	Obtenção de dados técnicos a respeito de aspectos ambientais e agrícolas.

Quadro 6 - Locais de visitas técnicas - Certificadoras.

Instituição	Local	Motivo/Contribuição
SGS	Rondonópolis Mato Grosso	Mapeamento inicial dos processos a respeito da classificação da fibra.
BMF	Rondonópolis Mato Grosso	Finalização do mapeamento dos processos a respeito da classificação da fibra.

Quadro 7 - Locais de visitas técnicas - Comerciais.

Instituição	Local	Motivo/Contribuição
COPERBEM	Campo Verde Mato Grosso	Identificação dos processos envolvendo os processos de beneficiamento.
Grupo Bom Jesus	Rondonópolis Mato Grosso	Reunião técnica com profissionais da área agrônômica e financeira sobre planejamento e gestão da cotonicultura.
Fazenda Bom Jesus	Pedra Preta Mato Grosso	Verificação dos processos a respeito do beneficiamento do algodão.
Tecelagem Bezerra de Menezes - TBM	Rondonópolis Mato Grosso	Verificação da transformação da pluma em produto comercial.
Martelli Transportes	Jaciara Mato Grosso	Entrevista e coleta de dados sobre logística terrestre aplicada à agricultura.
Yara Fertilizantes	Rondonópolis Mato Grosso	Visita técnica às instalações de fábrica de insumos.
Fazenda Torre	Rondonópolis Mato Grosso	Entrevista com pessoal técnico para coleta de dados a respeito do acompanhamento pós-plantio.
Industrial Bonfim de Exportação de Óleo – IBEL	Rondonópolis Mato Grosso	Visita técnica à fábrica de subprodutos derivados do algodão.

Quadro 8 - Locais de visitas técnicas - Governamentais.

Instituição	Local	Motivo/Contribuição
Instituto Nacional de Processamento de Embalagens Vazias-INPEV	Rondonópolis Mato Grosso	Visita técnica em unidade de reciclagem de embalagens plásticas de defensivos para mapeamento de processos.
PRÓ-SEMENTE	Rondonópolis Mato Grosso	Visita técnica em empresa de análise de qualidade de sementes.

Quadro 9 - Locais de visitas técnicas - Tecnológicas.

Instituição	Local	Motivo/Contribuição
TecSolo	Rondonópolis Mato Grosso	Verificação e identificação de processos e tipos de análises realizadas em solos para cultivares.
Unisystem	Rondonópolis Mato Grosso	Coleta de dados por meio de entrevistas sobre processos envolvendo plantio, acompanhamento e beneficiamento.
Fundação MT	Rondonópolis Mato Grosso	Visita com acompanhamento de pessoal técnico em cultivares, para verificação de metodologia para monitoramento de pragas na cotonicultura.

O ponto de partida para a formalização da ontologia é o mapeamento dos processos, seguido da identificação dos dados provenientes de cada processo e, finalmente, a etapa de formalização do modelo, contemplando o compartilhamento desse conhecimento. A ontologia constituída representa o conhecimento referente à cotonicultura e a formalização do mapeamento utilizado para construir a ontologia resulta em um padrão de metadados específico para esse segmento agrícola.

5.1 Delimitação das Classes de Dados

A formalização de ontologia é realizada, obedecendo às hierarquias dos elementos menos significativos para o mais significativo, conforme discutido por Urs (2007). A representação dos elementos, na agricultura, para esse domínio, não segue padrões *top-down* ou *bottom-up* de acordo com Inan (2010), mas adota uma hierarquia, conforme apresentado por Richetti et al (2003).

A Figura 15 mostra a sequência das etapas realizadas na pesquisa, dividindo-se em quatro níveis:

- Mapeamento dos processos: identificação dos principais processos utilizados na cotonicultura no Brasil com foco nos dados gerados em cada um deles.
- Formalização da ontologia: criação da ontologia, usando um modelo baseado em domínio para construção do conhecimento a respeito dos dados empregados na cotonicultura.
- Derivação do padrão de metadados: a criação do padrão de metadados consiste em formalizar o conhecimento obtido com a ontologia por meio de uma linguagem qualquer, sendo que, nesta pesquisa, será utilizado *XML*.
- Implementação da validação: a validação do padrão de metadados é realizada por meio de prova de conceito com duas implementações de *software*. A primeira, utilizando o padrão de metadados e a segunda, sem o uso desse padrão. Com isso, é possível verificar os benefícios dos padrões para homogeneização de dados.

Figura 15 - Visão geral do cenário abordado na pesquisa.

O uso da ontologia baseada em domínio possibilita a representação das classes chave e das classes de dados subordinadas hierarquicamente, em um cenário de investigação previamente definido. O processo de identificação das

classes segue métodos determinados em engenharia de *software*, de acordo com as características necessárias para a representação esperada na ontologia proposta.

A modelagem de dados computacionais, segundo Pressman (2009) e Sommerville (2010) pode ser representada por meio de sete modelos: modelo baseado em classes, modelo baseado em cenários, modelo comportamental, modelo de classe-responsabilidade-colaboração ou CRC, modelagem de dados, modelagem por especificação de controle e modelagem orientada a fluxo de dados. O uso de cada modelo de dados é condicionado ao segmento a ser representado e a pesquisa utilizou o modelo fundamentado em classes para formalização da ontologia. Essa escolha para o mapeamento das classes se justifica pela sua aderência às características que se pretende representar, possibilitando especificar: classes de dados, hierarquias e propriedades na cotonicultura.

O mapeamento dos principais processos foi realizado por meio de referências apresentadas nas pesquisas desenvolvidas por Richetti (2003) junto à EMBRAPA Algodão e Buainain (2007), em trabalho técnico compilado para o Ministério da Agricultura, Pecuária e Abastecimento – MAPA, em parceria com a Secretaria de Política Agrícola – SPA e Instituto Interamericano de Cooperação para a Agricultura – IICA. O detalhamento dos processos macros, descritos na Figura 16, são expostos nos modelos ilustrados nas Figuras 17, 18, 19 e 20.

5.2 Escopo de Classes abordadas

A delimitação das classes é realizada com a identificação dos processos envolvidos na cotonicultura. Nessa etapa, são construídos os fluxogramas para o mapeamento inicial de forma mais genérica dos principais processos. A Figura 16 apresenta uma visão geral de todos os processos mapeados, de acordo com a literatura técnica e dados obtidos complementarmente em visitas técnicas. Nesse contexto, foram identificadas cinco classes de dados chave (preparo do solo, análise de sementes, plantio, beneficiamento e comercialização) das quais derivam as demais classes subordinadas. As classes principais são comuns a outras culturas como soja e milho, sendo diferenciadas pelo tempo de realização de cada etapa, dependendo da cultura. O fluxograma é utilizado como referência para a identificação dos dados gerados nos processos, possibilitando a formalização da ontologia.

Figura 16 - Formalização de processos na cotonicultura com fluxogramas.

6. MAPEAMENTO DOS PROCESSOS

A contribuição do padrão de metadados para a cotonicultura, criado a partir da ontologia, consiste em melhorar a qualidade dos dados por meio de uma interface para padronização dos principais intercâmbios de dados entre processos ou pessoas. Grande parte dos problemas relacionados com intercâmbio e, conseqüentemente, com a qualidade de dados, deve-se ao empobrecimento dos dados como resultado de falhas de integridade, redundância, pouca completude e outras dimensões abordadas no capítulo 4. Os resultados abordados nesse capítulo estão relacionados com o mapeamento inicial das classes de dados.

6.1 Detalhamento dos Processos

O sistema produtivo é um conjunto de componentes interativos que tem como objetivo a produção de matérias-primas. No contexto da pesquisa, o conjunto de processos abordados na cotonicultura é referido por Castro (2000) como atividades “dentro da porteira da fazenda”, no qual os componentes se referem a ações, iniciando com a seleção de variedade e preparo do solo até o beneficiamento do algodão, caracterizado pela produção de pluma e comercialização de sementes. O padrão de metadados resultante concentra-se nos principais dados gerados na cotonicultura. O objetivo é definir a estrutura de dados hierarquicamente, com os vínculos mais relevantes entre os mesmos. Os processos relativos ao preparo do solo, abordados por Richetti (2003), foram utilizados como referência para a construção dos fluxos, ilustrados na Figura 17, em que o mapeamento dos processos tem início com a adubação e correção do solo. Essa Figura apresenta os processos macros na etapa do preparo do solo, quando foram mapeados quatro processos referentes à: coleta de amostras, preparo para análise, processamento de análises e correção de solo. Essa etapa da safra não é realizada de forma mandatória ou com o propósito de atender a regulamentações governamentais. As análises de solo representam uma etapa de total obrigação do responsável pelo cultivo, com o objetivo de identificar deficiências a serem corrigidas. Os processos se iniciam com amostras de solo enviadas a instituições governamentais ou particulares para realização de análises previamente definidas pelo pessoal técnico da área agrônômica; posteriormente, essas amostras são preparadas pela equipe técnica

para início das atividades de análises mecanizadas ou automatizadas, dependendo da solicitação. A etapa que segue se refere à emissão de laudos com informações para interpretação e início de eventuais ações corretivas na área investigada.

Figura 17 - Formalização dos processos referentes ao preparo do solo.

Os processos utilizados para a seleção da variedade mais adequada ao cultivo e ao controle de qualidade exigido por órgãos governamentais são expostos na Figura 18. A escolha da variedade de semente para uma região, feita com o auxílio de parâmetros referentes a características do solo e ambientais, pode determinar o sucesso ou o fracasso do cultivo. Nessa etapa, também são abordados processos responsáveis pelo gerenciamento de organismos geneticamente modificados. A autorização para o uso de sementes classificadas como OGM é uma

etapa importante no processo macro de análise de semente. Com a efetivação dos critérios de seleção e adequação das variedades a serem cultivadas, a fase seguinte refere-se ao plantio, sendo esse o processo final, identificado para a etapa de análise de sementes.

Figura 18 - Formalização dos processos referentes à análise de sementes.

A etapa do plantio determina a qualidade da matéria prima final e, para a fibra do algodão, processos como irrigação e aplicação de insumos são essenciais nessa fase da safra. A Figura 19 aborda os processos posteriores à escolha da variedade de semente a ser cultivada e à execução do plantio. O conjunto de processos mapeados nessa etapa é de suma importância, pois, além de envolver um período relativamente longo, abrange também várias etapas referentes ao acompanhamento das áreas cultivadas, demandando monitoramento constante para execução de ações preventivas ou corretivas, de acordo com características específicas de cada região da cultivar. A fase de acompanhamento do plantio possui processos pré-

definidos cronologicamente, iniciando com irrigações sazonais e aplicação de defensivos e insumos específicos, em datas previamente conhecidas pela equipe operacional responsável. Nessa etapa, também há grande demanda por recursos humanos tanto para operações manuais quanto ações com elevado grau de especialidade. Aspectos ambientais também são abordados em processos mapeados nessa etapa, envolvendo o procedimento de embalagens de defensivos com potencial risco para o meio ambiente, cujo descarte seguro é realizado em usinas de processamento de materiais em polos geograficamente estratégicos.

Figura 19 - Formalização dos processos referentes ao plantio.

Os últimos processos referem-se ao beneficiamento do algodão *in natura*, com início na colheita e concluindo com a criação dos sub-blocos ou fardinhos. Nessas etapas, os procedimentos envolvem a transformação final da matéria-prima e o controle de qualidade exigido por parâmetros internacionais. Inicia-se com a

logística de transporte terrestre do deslocamento do algodão bruto do local de colheita para as unidades beneficiadoras desse produto; posteriormente, características como peso e umidade são verificadas. Após o descaroçamento ou a separação entre a semente e a fibra do algodão, são realizadas duas análises referentes à classificação de cor e *high volume instrument* ou *HVI*, sendo a última, um conjunto de análises automatizadas para identificar características como: resistência da fibra, *Micronaire*, RD e outras necessárias para a comercialização da fibra do algodão.

Figura 20 - Formalização dos processos referentes ao beneficiamento.

6.2 Mapeamento das Classes de Dados

A partir da delimitação dos elementos iniciais envolvidos nesse cenário, foi possível realizar uma delimitação do escopo, resultando em um domínio de conhecimento específico. O cenário estudado na pesquisa refere-se à cotonicultura e, por isso, o ponto de partida para a construção da ontologia foi a área de cultivo e, posteriormente, a individualização dessa área em quadras ou subdivisões para plantio. Em seguida, foi identificado o primeiro subproduto, denominado sub-bloco ou fardão, representando uma unidade lógica com o produto bruto sem qualquer beneficiamento. Finalmente, determinou-se a unidade lógica final para o escopo estudado, denominado sub-bloco ou fardinho, que é a última unidade lógica contemplada na ontologia, pois os processos seguintes referem-se à comercialização e logística, e não são considerados na ontologia baseada em domínio, pelo fato de estarem fora do escopo proposto envolvendo a cotonicultura, conforme sugerido por Richetti et al (2003). A Figura 21 ilustra o escopo diretamente relacionado com a pesquisa, delimitando os processos relevantes da cotonicultura para extração dos conceitos e posterior mapeamento das classes de dados para construção da ontologia baseada em domínio. Apesar de existirem diversos processos identificados na cotonicultura, somente um reduzido grupo foi reconhecido como relevante para o domínio mais específico, referente à produção de fibra de algodão, de acordo com Richetti (2003) e Buainain (2007).

Figura 21 - Delimitação de domínio para extração de conceitos.

Fonte: Richetti (2003) e Buainain (2007).

A extração dos conceitos para construção da ontologia é realizada de forma genérica, a partir de um vocabulário informal. Em seguida, esses conceitos são transformados em classes para mapeamento dos membros subordinados ou

hierarquicamente inferiores. A identificação das concepções iniciais é realizada, a partir da compilação de referências bibliográficas e referências complementares, obtidas em reuniões técnicas para elencar os processos aspirantes a conceitos e, posteriormente, classes de dados em um segmento específico no escopo delimitado previamente. A cotonicultura, da mesma forma que outras culturas, como soja e milho, possui diversos processos e a escolha dos processos indicados para a representação de uma área específica depende, exclusivamente, do conhecimento que se pretende representar. Neste caso específico da ontologia de domínio, o conhecimento a ser exposto é restrito à produção de fibra de algodão, sendo esse o mais importante subproduto da cotonicultura.

Outro ponto imprescindível, quanto à escolha da ontologia baseada em domínio, refere-se à adequação desse modelo ao contexto da pesquisa. Estas características: verticalização de escopo, especificação de classes singulares e especificação de atributos singulares, são essenciais para a construção precisa desse tipo de conhecimento. A exemplo disso pode-se utilizar o conceito: área de cultivo com o seguinte fragmento:

Área: **Agricultura**

Escopo: Cotonicultura [**Domínio: Produção de Fibras de Algodão**]

Elemento Informal: Área de cultivo

Conceito: **Fazenda**

NomeFazenda [**Domínio Alfanumérico**]

ÁreaTotal [**Domínio Numérico**]

Quadra

NomeQuadra [**Domínio Alfanumérico**]

ÁreaParcial [**Domínio Numérico**]

O fragmento do modelo de classes de dados ilustra a definição de escopo referente à produção de fibra na cotonicultura dentro da área macro Agricultura. Elucida, também, a definição de escopo para alguns atributos das classes fazenda e quadra, especificando regras de permissibilidade de valores. Outro ponto importante refere-se à singularidade de nomenclaturas utilizadas tanto para definir classes derivadas de conceitos quanto seus atributos.

O mapeamento das classes de dados tem o objetivo de identificar os requisitos necessários para a criação do modelo que represente o conhecimento

desse segmento, fornecendo melhor compreensão dos processos globais. A seguir, será apresentado o detalhamento de cada classe de dados; o padrão de metadados proposto utiliza as classes de dados mapeadas como referência para definição da estrutura de *tags*.

Quadro 10. Classes de dados - preparo de solo.

Classe	Função
Fazenda (Farm)	é a principal classe da qual todas as demais são derivadas e esse contexto representa a classe de dados mais visível no setor agrícola, do ponto de vista administrativo e gerencial.
Quadra (Square)	para atender às particularidades da área de cultivo são realizadas divisões proporcionais, denominadas quadras ou talhões. No início de cada safra, é necessário considerar o preparo do solo e isso consiste em quaisquer correções efetuadas com insumos, tais como cal e gesso.
Análise de Solo (SoilAnalysis)	segundo Buainain (2007), análises de solos são necessárias para verificação da necessidade de correções antes do plantio. A classe SoilAnalysis estabelece as estruturas para armazenamento dos dados referentes a análises e controle de qualidade do solo nas áreas de cultivo.
Solo (Soil)	análises básicas relacionadas com a qualidade do solo a ser cultivado.
Fertilizante (Fertilization)	dados referentes à fertilização do solo.
Fertilizantes Micros (FertilizationMicros)	dados referentes a micronutrientes presentes no solo.
Fósforo (Phosphorus)	dados referentes ao fósforo presente no solo.
Calcário (LimeStone)	dados quantitativos referentes ao calcário presente no solo.
Folha (Leaf)	análise foliar do solo
Folha Macro (LeafMacro)	análise foliar macro, de macros nutrientes do solo.
Folha Micro (LeafMicro)	análise foliar de micros nutrientes do solo.
Sal Mineral (MineralSalt)	dados a respeito do potencial de sal mineral existente no solo.
Análise de Ração (Feed)	Dados a respeito de resíduos orgânicos existentes no solo.
Resíduo Orgânico (OrganicResidue)	análise de resíduos orgânicos do solo.
Resíduo Orgânico Macro (OrganicResidueMacro)	análises de resíduos orgânicos macros do solo.
Resíduo Orgânico Micro (OrganicResidueMicro)	análises de micros resíduos orgânicos do solo.
Resíduos (Residues)	análise residual de características do solo.

Figura 22 - Classe de dados referentes ao preparo de solo.

Quadro 11. Classes de dados – análise de sementes.

Classe	Função
Cultivo (Cultivation)	definição dos dados para o cultivo como período e cultivares.
Variedade (Variety)	diversidade da cultura selecionada para o plantio.
Análise de Sementes (SeedAnalysis)	resultados das análises de sementes para controle de qualidade na fase de pré-plantio.
Insumos (Consumables)	dados a respeito da aplicação de insumos utilizados no pré-plantio.
Gesso (ConsumablesGypsum)	dados a respeito da aplicação de gesso no pré-plantio.
Fertilizantes (ConsumablesFertilizer)	dados a respeito da aplicação de fertilizantes diversos no pré-plantio.
Calcário (ConsumablesLimestone)	dados a respeito da aplicação de calcário no pré-plantio.

Figura 23 - Classe de dados referentes à análise de sementes.

Quadro 12. Classes de dados - plantio.

Classe	Função
Defensivos (Pesticides)	defensivos utilizados no acompanhamento da safra.
Descarte de Embalagens Vazias (DiscardEmptyPackaging)	dados referentes ao processamento de embalagens de defensivos após aplicações.
Tipos de Embalagens (TypePacking)	define os tipos de embalagens utilizadas em defensivos, de acordo com classificação fornecida pelo Instituto Nacional de Embalagens Vazias.
Não Lavável (NonWhasable)	define os tipos de embalagens não laváveis para defensivos.
Flexíveis (Flexible)	define os tipos de embalagens flexíveis para defensivos.
Não Flexível (NonFlexible)	define os tipos de embalagens não flexíveis para defensivos.
Secundária (Secondary)	define os tipos de embalagens secundárias para defensivos.
Lavável (Whasable)	define os tipos de embalagens laváveis para defensivos.
Fungicida (Fungicide)	fungicidas aplicados no acompanhamento do plantio.
Herbicida (Herbicide)	herbicidas aplicados no acompanhamento do plantio.
Inseticida (Insecticide)	inseticidas aplicados no acompanhamento do plantio.
Colheita (Harvest)	dados relacionados com a formalização de informações referentes à colheita, como o período de execução.
Unidade de Processamento de Algodão (CottonProcessingUnit)	dados referentes à unidade de processamento de algodão responsável pelo beneficiamento ou a transformação de algodão bruto em subprodutos comerciais.
Recursos Hídricos (WaterResources)	representação dos dados a respeito de recursos hídricos utilizados nas quadras, no acompanhamento da safra.
Análise de Água (WaterAnalysis)	análises de características da água utilizada na irrigação das áreas cultivadas.
Bloco Externo (ExternBlock)	dados referentes à matéria-prima de terceiros; neste contexto, algodão bruto para beneficiamento.

Figura 24 - Classe de dados referentes ao plantio.

Quadro 13 - Classes de dados - beneficiamento.

Classe	Função
Blocos (Blocks):	a atividade final no campo inclui a colheita do algodão, seguida da compactação em fardos de aproximadamente 12 toneladas também chamados de fardões. Os fardos de algodão são, temporariamente, armazenados no próprio local de colheita até o momento de serem removidos para a unidade de beneficiamento de algodão. Ainda nessa etapa, a análise de umidade é realizada para verificar a qualidade do algodão a ser beneficiado, com o objetivo de utilizar essa análise como referência para controlar a qualidade do algodão após o processo de descaroçamento, conforme relatório técnico apresentado pela Ampasul (2011).
Sub-blocos (SmallBlocos):	de acordo com Richetti et al (2003), o beneficiamento da pluma é finalizado com o empacotamento das fibras descaroçadas, em cuja etapa as fibras são divididas em pequenos blocos, denominados fardinhos.
HVI:	dados a respeito de análises mecanizadas, realizadas para obtenção de variáveis de controle de qualidade, denominada <i>High Volume Instrument</i> (RILEY, 1997).
Processamento da Amêndoa (NutProcessing):	dados resultantes do processamento de derivados da amêndoa do algodão.
Processamento da Casca (PeelProcessing):	dados decorrentes do processamento de derivados da casca do algodão.
Processamento da Semente (SeedProcessingUnit):	dados a respeito do processamento de sementes para novos cultivos.
Classificação Visual (VisualClassification):	Classificação das características visuais da fibra do algodão, como a coloração.
Documento de Transporte (DocTransport):	documentos obrigatórios para o transporte de produtos em sua fase bruta ou beneficiada.

Figura 25 - Classe de dados referentes ao beneficiamento.

Comercialização (Commercialization): dados de destino final para o algodão processado e referem-se a informações comerciais com cliente, quantidades comercializadas, portos de embarque, entre outros.

O uso da ontologia baseada em domínio foi adequado em razão destas características: especificação vertical de escopo, especificação de classes singulares e especificação de atributos singulares.

6.3 Padrão de Metadados Proposto

O padrão de metadados derivado da ontologia para cotonicultura pode contribuir, no intercâmbio de dados, com a criação de interfaces em camadas intermediárias. A formalização lógica do padrão é realizada com a linguagem *XML*, cuja escolha se justifica pela versatilidade, tanto na representação quanto na transparência multiplataforma. Na Figura 26 é ilustrado, graficamente, as classes de dados mapeadas e descritas previamente e, nessa representação, é possível verificar todas as classes principais com os seus nós subordinados, vinculados a classes de dados derivadas. A representação completa das classes com todos os atributos é mostrada na Figura 27, no qual são ilustradas, além da hierarquia de classes, o grau de subordinação e todas as suas características.

Figura 26 - Classes de dados propostas.

Figura 27 - Resumo das classes de dados com atributos.

A validação do padrão pode ser realizada de duas formas: a primeira, por meio do monitoramento sistemático de soluções implementadas com o padrão de metadados, proposto por meio de variáveis previamente definidas em função do tempo e a segunda forma por intermédio de uma prova de conceito; neste caso, é necessário criar um cenário tecnicamente adequado para a implementação de protótipos, utilizando o padrão sugerido. Isso possibilita análises referentes ao comportamento das aplicações empregando o padrão de metadados, confrontando-

as com implementações similares sem o uso do padrão e verificando os resultados obtidos nas duas situações. A validação do padrão proposto será atingida por meio de prova de conceito, discutida no capítulo 4, em razão da viabilidade técnica quanto à realização de testes e simulações.

7. PROVA DE CONCEITO

A prova de conceito está dividida em duas partes: análise do padrão de metadados comparado a outros padrões de armazenamento de dados e o desenvolvimento de *softwares* em dois diferentes paradigmas de desenvolvimento de *software* (arquitetura orientada a serviços e *cliente/server*) utilizando a estrutura de metadados. A análise de bases de dados em 13 empresas agrícolas mostrou uma média de 10 GB de dados por empresa, com uma variação entre 2 e 50 GB. Nesse contexto, o padrão de metadados é derivado a partir das classes de dados formalizados na ontologia, e a relação entre essas classes de dados na ontologia e *tags* disponíveis no padrão XML resultante é de um para um, pois todas as classes possuem uma *tag* correspondente com suas respectivas subclasses e atributos. O documento XML serve como referência para criação de estruturas de armazenamento de dados em bancos de dados, desenvolvimento de interfaces em aplicações distribuídas e interfaces WSDL em *webservices*. Outra aplicação importante para o padrão de metadados está no uso como referência na homogeneização de dados para integração de bases de dados heterogêneas em aplicações como: *BI*, *KDD* e *data mining*. O padrão de metadados para cotonicultura contribui para a criação de uma interface comum de intercâmbio de dados, uma vez que metadados fornecem independência de ferramentas como banco de dados.

7.1 Qualidade de Dados Aplicada à Agricultura

A qualidade dos dados tem consequências diretas na geração de informações de valor e eficiência no gerenciamento organizacional. Segundo McGilvray (2008), dados podem ser categorizados ou agrupados de acordo com características comuns e entender o relacionamento e dependência entre as diferentes categorias de dados contribui para o gerenciamento mais eficiente. De acordo com Wang (2002), diversas iniciativas corporativas importantes, como *business-to-business commerce*, *integrated supply chain management* e *enterprise resource planning*, estão sujeitas a falhas relacionadas ao gerenciamento de dados ineficientes. Ainda segundo Wang (2002), 70% dos dados utilizados em aplicações corporativas apresentam algum tipo de erro ou falha. Por isso, grandes corporações têm investido tempo para reduzir *gaps* em dados gerados ou utilizados nos mais diversos segmentos, como indústria, prestação de serviços e agricultura. Parte das pesquisas

tem o objetivo de solucionar problemas e envolvem o conceito de enriquecimento de dados ou melhoria de dados existentes, por meio da agregação de detalhes complementares a dados conhecidos e consolidados.

Na agricultura, a qualidade dos dados também é decisiva para manter e criar formas de gerenciar elementos confiáveis e, em alguns casos, torna-se difícil aplicar conceitos de enriquecimento de dados, considerando a pluralidade das fontes utilizadas nos processos. A ausência de um padrão de dados específicos propicia consequências indesejadas, como duplicidades de dados e inconsistências. Na Figura 28 é ilustrado um cenário de duplicidade e imprecisão de dados em uma situação real, mostrando as diferentes visões que variados setores têm a respeito do mesmo dado, no contexto da área para cultivo, com 500 hectares físicos, cujo setor de planejamento trabalha com valor quantitativo de 510 hectares, prevendo perdas de insumos na execução do plantio e, para evitar atrasos nesse processo, atribui a margem de erro de 10 hectares a mais.

Os recursos humanos necessários para o cultivo são mensurados de acordo com a cultura a ser plantada e essa métrica reflete diretamente na quantidade de pessoas necessárias por hectare, considerando a ausência eventual de trabalhadores por motivos diversos. Devido a isso, o planejamento considera 30 hectares a mais. A aquisição de insumos por meio do setor de compras prevê uma margem de erro habitual com base no histórico de plantio e, dessa forma, não deve estar abaixo do nível mínimo ou poderão ocorrer problemas na execução do plantio. Dessa forma, o planejamento é realizado, utilizando uma margem de 20 hectares acima da quantidade real. Finalmente, o setor técnico ou planejamento técnico feito por engenheiros agrônomos usa a margem de 05 hectares a mais da quantidade real, prevendo eventuais ajustes na área de plantio.

de metadados para estruturação no intercâmbio de dados, no segmento agrícola abordado na pesquisa.

7.2.1 Escopo da Prova de Conceito

O objetivo da prova de conceito, nesse cenário, é apresentar, de forma prática, o uso do padrão de metadados derivado da ontologia, no desenvolvimento de *softwares*, contemplando aspectos como intercâmbio de dados e qualidade dos mesmos na cotonicultura e destacando, também, a melhoria no processo de desenvolvimento de *software* em aplicações computacionais com o uso do padrão de metadados proposto.

7.2.2 Ambiente para Testes

A realização dos testes depende de ambiente computacional adequado para atender aos requisitos a serem investigados. Esse ambiente é materializado em *softwares* e *hardwares* mandatórios na geração e transporte de dados, nos testes em que serão submetidos os protótipos e, posteriormente, no tratamento e análise dos dados gerados.

Quadro 14 - Lista de requisitos para prova de conceito.

Recurso	Tipo	Fornecedor	Descrição
<i>Apache Web Server</i>	<i>Software</i>	<i>Apache Foundation</i>	Servidor <i>web</i> destinado a hospedar os <i>webservices</i> desenvolvidos em <i>PHP</i>
<i>Borland Delphi Studio</i>	<i>Software</i>	<i>Embarcadero</i>	Compilador utilizado para o desenvolvimento dos protótipos em ambiente baseado em janelas
<i>Data Modeler</i>	<i>Software</i>	<i>Oracle Corporation</i>	Ferramenta <i>CASE</i> para criação do modelo lógico das tabelas e relacionamentos
<i>OS Windows 7</i>	<i>Software</i>	<i>Microsoft</i>	Plataforma operacional utilizada para hospedagem do SGBD, <i>web service</i> e protótipos
<i>PHP</i>	<i>Software</i>	<i>The PHP Group</i>	Linguagem de programação <i>script</i> utilizada para criação dos <i>webservices</i>
<i>PsPad</i>	<i>Software</i>	<i>PsPad</i>	Editor de <i>scripts</i> multilinguagem para criação dos <i>scripts</i> em <i>PHP</i>
<i>Oracle 10g Xe</i>	<i>Software</i>	<i>Oracle Corporation</i>	Sistema gerenciador de banco de dados.

7.2.3 Mapeamento entre Modelos de Dados

A análise comparativa entre os diferentes formatos de armazenamento de dados partiu da formalização de um fragmento da ontologia em classes de dados e, para isso, foi realizado o mapeamento entre um fragmento do modelo ExR referente à rastreabilidade com as classes de dados correspondentes.

Figura 29 - Fragmento do mapeamento entre modelo ExR e classes de dados.

A partir do mapeamento das classes de dados correspondentes às tabelas do modelo ExR utilizado como referência, foi construída uma estrutura para armazenamento de dados a serem comparados em três formatos: *comma-separated values* ou CSV, dicionário de dados e o padrão de metadados proposto, conforme mostrado na Figura 30.

Figura 30 - Modelo ExR correspondente à estrutura de *tags*.

Após a criação do modelo ExR com a estrutura proposta no padrão de metadados, foi usado um grupo de tabelas correspondente às classes fazenda (*farm*) e quadras (*square*), conforme mostrado na Figura 31.

Figura 31 - Fragmento do modelo ExR.

O fragmento do modelo ExR, desenvolvido para implementação dos protótipos, atende a requisitos mínimos de modelagem de dados aplicados a banco de dados como: definição chave primária (Farm:IDFarm, Square:IDSquare; Square:IDFarm), definição de grau de cardinalidade entre tabelas

1:N (cada fazenda, por ter várias quadras associadas) e domínios de dados definidos com clareza (integer, varchar2, float e number). Esses requisitos referem-se a regras básicas de modelagem de dados para prevenção de inconsistências e redundância de dados.

A estrutura de dados, referente ao padrão de metadados, criada em um gerenciador de banco de dados, possibilita a comparação nos três formatos mencionados e analisados no item 4.2.5. A Figura 32 ilustra uma área de plantio (*farm*) com as divisões de quadras (*square*), cujo cenário foi criado a partir de uma área física existente, com alterações nas dimensões das quadras, e seu objetivo é utilizar dados com características reais para a verificação do comportamento do padrão de metadados proposto nesse cenário.

Figura 32 - Mapa ilustrativo de divisão de quadras em fazendas.

Os valores gerados para a construção do mapa serão utilizados para o preenchimento da infraestrutura de tabelas utilizada pelos dois protótipos desenvolvidos para prova de conceito. Apesar de não serem valores reais uma proximidade direta entre os valores apresentados e valores das quadras reais.

7.2.4 Infraestrutura

A comparação do padrão de metadados com outros padrões de armazenamento de dados é realizada, utilizando métricas que envolvem a estrutura de armazenamento e melhorias na recuperação e visualização dos dados. As aplicações também estão divididas em dois tipos de implementação: com metadados e sem metadados. A proposta é analisar a contribuição do padrão de metadados nesse contexto, tanto na melhoria do intercâmbio de dados quanto no processo de desenvolvimento de *software*.

Figura 33 - Estrutura geral da prova de conceito.

A criação dos protótipos nesses paradigmas de programação partiu de um ponto comum: a construção da infraestrutura dos dados com os respectivos relacionamentos. Conforme ilustrado na Figura 34, essa infraestrutura foi formalizada em um modelo Entidade-Relacionamento, ou simplesmente ExR, empregando um sistema de rastreabilidade como referência. O uso de uma infraestrutura lógica para rastreabilidade aborda classes de dados semelhantes às mostradas no padrão de metadados, como: fazendas, quadras, blocos e sub-blocos. As demais classes expostas no modelo ilustrado na Figura 34 são utilizadas como complementos para fornecer recursos de localização geográfica, certificação e controle de qualidade.

meio de *webservices*. Nesse exemplo, foram adotadas duas interfaces para acesso aos serviços: a primeira, em ambiente gráfico e a segunda, em ambiente textual, conforme ilustrado na Figura 37. O objetivo é mostrar a independência de interface para esse modelo de aplicação. Nesse contexto, a principal característica analisada no protótipo, em utilizando metadados, é o formato do retorno das requisições nas consultas, que pode ser disponibilizado em dois formatos: proprietário, com estrutura definida de acordo com critérios do desenvolvedor da solução e interface com *Web Services Description Language* ou *WSDL*, padrão para criação de interface usada por *webservices*. O formato adotado foi o primeiro, aproveitando como interface o padrão de metadados proposto. Dessa forma, a customização e o tratamento do retorno das requisições foram realizados por meio da interpretação de dados em *XML*, com *tags* criadas a partir das classes de dados apresentadas no padrão de metadados proposto.

Figura 36 - Modelo BPEL para criação de serviços para cotonicultura.

Os protótipos criados para os testes, utilizando metadados, possuem duas interfaces para visualizações dos dados em ambientes distintos: a primeira interface apresenta os dados de forma textual em uma consulta simples; a segunda interface, desenvolvida em ambiente gráfico, aponta um *treeview* para visualização dos dados, e esse tipo é instituído por meio de elemento gráfico, expondo uma visão hierárquica das informações, em que, cada item, em geral chamado de nó, pode ter diversos subitens. Os componentes podem ser expandidos até a apresentação de todos os subitens. *Treeviews*, geralmente, são utilizados para facilitar a navegação entre dados, como, por exemplo, um documento *XML*.

Figura 37 - Interfaces dos protótipos desenvolvidos.

A segunda implementação foi criada sem o padrão de metadados, utilizando recursos proprietários de ferramentas específicas e a escolha desse paradigma justifica-se pela adoção, em larga escala, no desenvolvimento de *software* para o agronegócio. Parte dessa preferência deve-se à simplicidade da arquitetura em geral, utilizando duas camadas para o desenvolvimento de *softwares*. Na Figura 38 é ilustrada a interface desenvolvida para o gerenciamento de quadras em fazendas.

Figura 38 - Aplicação com arquitetura proprietária.

Descrição	Área
02	189.58
03	224.26
01	111.06
04	49.19
05	273.78
06	123.35
07	187.06
08	196.21
09A	132.93
09B	85.73
10	238.56
11	274.15
12	268.2
13	136.54

Os dois protótipos foram criados com ênfase na recuperação dos dados referentes a fazendas (*farm*) e quadras (*square*), sendo o primeiro desenvolvido com uma linguagem de programação em ambiente *desktop* e o segundo em ambiente *web*.

7.2.5 Análise dos resultados

As análises comparativas são apresentadas em três etapas: a primeira, relacionada com o desenvolvimento de software; a segunda analisa o comportamento do intercâmbio de dados nos dois paradigmas supracitados e a terceira verifica o comportamento dos dados em três padrões de formalização de dados.

7.2.5.1 Métricas de Desenvolvimento de Software

As análises relacionadas com o desenvolvimento de *software* apresentaram melhores resultados com a implementação utilizando o padrão de metadados, em relação à implementação sem o padrão de metadados. Os primeiros resultados foram identificados no processo de codificação da aplicação envolvendo a complexidade das estruturas utilizadas. A análise referente ao desenvolvimento de *software* foi realizada empregando três métricas: complexidade ciclomática, linhas

de código e número de membros aplicados à implementação de procedimentos e funções criadas para a composição do código.

A métrica complexidade ciclomática proposta por McCabe (1976), ilustrado no Gráfico 5, avalia a quantidade de funções de controle de fluxo e condicionais (*if..then..else*, *while*, *repeat*) encontradas no código. A quantidade excessiva de condicionais podem indicar pouca estruturação lógica no projeto do *software*; o protótipo, utilizando metadados, apresentou relativa redução na quantidade de condicionais. Segundo Jones (2012), a métrica de linhas de código (LOC) foi exposta na década de 1960, com o propósito de aferir o valor econômico de projetos de desenvolvimento de *software*, a partir da quantidade de linhas de código demandada para a sua construção. Os Gráficos 6 e 7 ilustram a quantidade total de linhas e a quantidade total de funções utilizadas para a criação dos dois protótipos de *software* nos dois paradigmas. A consequência da ausência de componentes de interface para a implementação, utilizando metadados, reduz a quantidade de linhas usadas no código fonte.

Gráfico 5 - Complexidade ciclomática.

Gráfico 6 - Quantidade de linhas de código.

Gráfico 7 - Número de membros.

Nos Gráficos 8, 9 e 10 são ilustrados o comportamento numérico da implementação de códigos em funções internas, cujas métricas avaliam, especificamente, a codificação de funções e procedimentos utilizados para redução de processamentos e cálculos frequentes. O Gráfico 8 ilustra a quantidade média de caracteres usada para implementação em cada linha. O excesso de caracteres por linha aponta para um grau de complexidade indesejado para codificações que poderiam ser escritas de forma mais simples.

Gráfico 8 - Quantidade média de caracteres por linha.

A quantidade de parâmetros de entrada e saída, definidos em funções e procedimentos, pode determinar o grau de complexidade utilizado para a construção do código fonte; funções com excesso de parâmetros de entrada e saída dificultam a manutenção e extensibilidade das mesmas, principalmente, quando estão vinculadas a classes. As implementações usadas como referência para análise apresentaram o mesmo comportamento nas duas situações quanto a esse quesito.

Gráfico 9 - Quantidade média de parâmetros de entrada e saída.

Funções são úteis para a redução de linhas códigos e os valores das métricas apontadas no Gráfico 10 indicam uma melhora significativa quanto ao uso de funções em aplicações implementadas com metadados. Nesse quesito, o maior número de funções instaladas reflete em maior reuso de código e ganho em tempo na construção de códigos.

Gráfico 10 - Uso de funções.

Deve-se ressaltar que metadados contribuem para a criação de interfaces de comunicação de dados entre *softwares*; a codificação de aplicativos assume características diferentes de acordo com o paradigma de desenvolvimento adotado. Aplicações, contendo características utilizadas para o incremento do protótipo com o

padrão de metadados apresentado na pesquisa, em geral, precisam de regras para definir a interface para consumir os serviços disponibilizados. A estrutura usada para desenvolvimento do protótipo sem o uso de metadados emprega padrões de troca de mensagem proprietários; dessa forma, não é mandatório o uso de um padrão explícito para intercâmbio de dados.

7.2.5.2 Métricas de Intercâmbio de Dados

Segundo McGilvray (2008), dados podem ser categorizados ou agrupados de acordo com características em comum. A definição inicial das dimensões para qualificação de dados foi realizada a partir das referências de Wang et al. (2002) e McGilvray (2008) as quais apresentam abordagens complementares nessas indicações. Métricas podem assumir importâncias diferentes, dependendo do contexto, e os critérios para a avaliação de dados também podem admitir caráter subjetivo. Os requisitos para análise dos dados, nos processos de intercâmbio, foram elaborados utilizando 19 dimensões mapeadas em 20 referências técnicas sobre qualidade de dados com faixa de valores entre zero e quatro, representando o número de consultores participantes da validação das dimensões aplicadas à cotonicultura. Posteriormente, foi verificada a proporcionalidade para cada dimensão, definindo o grau de importância para cada uma. Apesar de as métricas mapeadas em referências literárias, apresentadas no Gráfico 11, mostrarem um grau de recorrência diferente, todas se revelam importantes, de acordo com os autores consultados.

Gráfico 11 - Análise das dimensões de intercâmbio de dados.

Algumas métricas para análise da prova de conceito não foram utilizadas na pesquisa por não possibilitarem, de forma crítica, comparações ao uso de metadados em confronto com outros padrões. Por isso, neste contexto, foram adotadas: acessibilidade, completude, conformidade, especificação de dados, facilidade de uso e manutenibilidade, fonte, objetividade, qualidade de apresentação, transatabilidade e validade. Não foram usadas: acurácia, cobertura de dados, consistência, duplicações, empobrecimento de dados, integridade, intemporal, relevância e segurança de acesso. A seguir, são apresentadas as dimensões com dados, sem o mapeamento com o padrão de metadados proposto, cujos dados são confrontados com o uso do padrão para a verificação da eficiência do seu uso nesse contexto. As dimensões selecionadas foram submetidas a um conjunto de dados referente a um fragmento da base de dados.

O conjunto de dados a seguir será formalizado nas dez dimensões selecionadas para validação do padrão de metadados verificando a eficiência da estrutura com esse fragmento de dados.

Dados Originais: Colorado, 1, 1, T-1, 92, -16, 083754, -54, 930124
--

Quadro 15 – Dimensão Acessibilidade.

Contribuição	Disponibilização dos dados de forma fácil para o usuário ou consumidor de dados. O usuário de dados tem informações prévias a respeito do acesso aos dados. A característica hierárquica da estrutura dos dados proporciona facilidade quanto a localização de um dado específico.
<pre> FARM ├── FARMNAME:Colorado ├── SQUARE │ ├── NAME:T-1 │ ├── AREA:92 │ ├── LATITUDE:-16.083754 │ └── LONGITUDE:-54.930124 </pre>	

Quadro 16 – Dimensão Completude.

Contribuição	Medida referente à disponibilidade do dado em atender às necessidades de uso mesmo incompleto. Mesmo dados incompletos são compreensíveis em razão do uso de <i>tags</i> para identificação.
<pre> FARM ├── FARMNAME:Colorado ├── SQUARE │ ├── NAME:T-1 │ ├── AREA:92 │ ├── LATITUDE:-16.083754 │ └── LONGITUDE:-54.930124 </pre>	

Quadro 17 – Dimensão Conformidade.

Dimensão	Conformidade
Contribuição	Descreve como os dados aderem aos padrões existentes e como estão representados em um formato esperado. A relação entre as classes de dados e os dados a serem representados devem ser 1:1 (um-para-um). Dados sem classes para representa-los indicam que a estrutura do padrão está deficiente.
FARM FARMNAME: Colorado SQUARE NAME: T-1 AREA: 92 LATITUDE: -16.083754 LONGITUDE: -54.930124	

Quadro 18 – Dimensão Especificação de Dados.

Contribuição	Determina existência, plenitude, qualidade e documentação de padrões de dados, modelos de dados, regras de negócios, metadados e referências de dados. Um ponto importante em um padrão de metadados é a possibilidade de definir domínios para representação de dados, dessa forma, dados representados tipicamente por valores numéricos como área, latitude ou longitude não devem oferecer recursos de representação para outros tipos de valores como letras ou símbolos especiais.
FARM FARMNAME: Colorado SQUARE NAME: T-1 AREA: 92 LATITUDE: -16.083754 LONGITUDE: -54.930124	

Quadro 19 – Dimensão Facilidade de Uso e Manutenabilidade.

Contribuição	Refere-se ao grau desejado para o acesso e uso do dado e também ao grau no qual pode ser atualizado e gerenciado. O padrão deve oferecer recursos para localização e manutenção de dados específicos.
FARM <ul style="list-style-type: none"> -FARMNAME:Colorado -SQUARE <ul style="list-style-type: none"> -NAME:T-1 -AREA:92 -LATITUDE:-16.083754 -LONGITUDE:-54.930124 	

Quadro 20 – Dimensão Fonte.

Dimensão	Fonte
Contribuição	Possibilidade de descrever e identificar a origem dos dados.
FARM <ul style="list-style-type: none"> -FARMNAME:Colorado -SQUARE <ul style="list-style-type: none"> -NAME:T-1 -AREA:92 -LATITUDE:-16.083754 -LONGITUDE:-54.930124 -Documentation:source=URI reference xml:lang=language 	

Quadro 21 – Dimensão Objetividade.

Contribuição	Os dados devem ser objetivos e não depender de julgamento, interpretação ou avaliação. As <i>tags</i> devem representar singularmente cada dados dispensando a necessidade de julgamento quanto ao seu significado.
FARM <ul style="list-style-type: none"> -FARMNAME:Colorado -SQUARE <ul style="list-style-type: none"> -NAME:T-1 -AREA:92 -LATITUDE:-16.083754 -LONGITUDE:-54.930124 	

Quadro 22 – Dimensão Apresentação.

Contribuição	Qualidades de apresentação dos dados aos destinatários; nesse aspecto, estão envolvidos formato e aparência, relacionados com a informação resultante. Apesar do padrão de metadados especificar inicialmente infraestrutura também é esperado sejam fornecidos recursos para apresentação de dados, isso é obtido com uma especificação de dados clara e concisa das classes de dados no domínio abordado.
FARM FARMNAME: Colorado SQUARE NAME: T-1 AREA: 92 LATITUDE: -16.083754 LONGITUDE: -54.930124	

Quadro 23 – Dimensão Transabilidade.

Contribuição	Medida para verificar o grau desejado para o dado, em transações entre processos ou resultados. A troca de dados entre processos organizacionais ou corporativos é fundamental para o intercâmbio de informações e geração de conhecimento, quando necessário.
FARM FARMNAME:Colorado SQUARE NAME:T-1 AREA:92 LATITUDE:-16.083754 LONGITUDE:-54.930124	

Os dados formalizados no padrão de metadados apresentam uma estrutura simples, sem especificações adicionais quanto à composição da estrutura a ser transferida entre processos; dessa forma, o uso para intercâmbio de dados é um ponto importante.

Quadro 24 – Dimensão Validade.

Contribuição	Determina o grau necessário para dados ou informações a serem mensurados. Medida complementar a especificação de dados, o requisito básico está na definição de limites para cada representação de dados. <i>Tags</i> como latitude além de representar valores numéricos deve fornecer recursos para a definição de limites mínimos e máximos resultando em um dado válido.
<pre> FARM ├── FARMNAME: Colorado └── SQUARE ├── NAME: T-1 ├── AREA: 92 ├── LATITUDE: -16.083754 └── LONGITUDE: -54.930124 </pre>	

A justificativa para algumas dimensões de dados serem excluídas da análise deve-se ao fato de o padrão de metadados não possuir características que proporcionem melhorias aos dados em um contexto específico; dimensões relacionadas com precisão e medida não são beneficiadas com padrões de metadados. Nesse contexto, dados numéricos ou alfanuméricos são resultantes da execução de funções e procedimentos que devolvem valores estáticos. A sua formatação junto a um padrão não melhora o seu valor, pois o padrão fornece somente recursos para armazenamento ou apresentação de dados. A seguir, são apresentadas algumas ilustrações com as dimensões excluídas da análise:

Medida	Acurácia
Descrição	Refere-se à medida de correção do conteúdo dos dados.
Justificativa	O uso de padrões de metadados não melhora a precisão ou corrige eventuais dados imprecisos; metadados somente reestruturam os dados.
Dados	-16.083754
Metadados	<LATITUDE>-16.083754</LATITUDE>
Medida	Cobertura de Dados
Descrição	Utilizada para mensurar a disponibilidade e abrangência dos dados no universo no qual representa.
Justificativa	Padrões de metadados formalizam conjuntos de dados disponíveis em um contexto; a ausência de dados para melhorar a descrição de um contexto refletirá diretamente na representação resultante do uso de metadados.
Dados	Colorado, São Lourenço

Metadados	<FARM> <FARMNAME>Colorado</FARMNAME> <SOIL> </SOIL> <WATERRESOURCES>São Lourenço</WATERRESOURCES> </FARM>
Medida	Consistência
Descrição	Equivalência da informação usada em diversos bancos de dados, aplicativos, sistemas e processos para criar dados equivalentes.
Justificativa	O padrão de metadados não fornece recursos para homogeneizar dados provenientes de fontes diferentes.
Dados	DB1: área=92 hectares. DB2: área=38.01 alqueires.
Metadados	<AREA>92</AREA> ou <AREA>38.01</AREA> não há medida de equivalência.
Medida	Duplicações
Descrição	Característica não desejada entre sistemas para qualquer tipo de dado, registro de dados ou mesmo um conjunto de dados.
Justificativa	O padrão de metadados formalizará conjuntos de dados mesmo que estejam duplicados; não há regras objetivas para eliminar eventuais duplicações.
Dados	-16.083754 -16.083754
Metadados	<LATITUDE>-16.083754</LATITUDE> <LATITUDE>-16.083754</LATITUDE>
Medida	Empobrecimento de Dados
Descrição	Verifica a taxa de utilidade do dado ou a diminuição da sua importância.
Justificativa	Metadados não permitem monitorar a qualidade do dado, verificando quando o mesmo precisa ser atualizado para que tenha relevância ou valor.
Dados	2014/2014, T-1, 92
Metadados	<SEASON>2014/2014<SEASON> <SQUARE>T-1<SQUARE> <AREA>92<AREA>
Medida	Integridade
Descrição	Refere-se à existência, validade, estrutura, conteúdo e outras características básicas do dado; inclui medidas fundamentais da qualidade de dados como taxa de preenchimento e faixas de valores permissíveis e integridade referencial.
Justificativa	Algumas medidas como: integridade referencial e taxa de preenchimento são controladas em ferramentas específicas de gerenciamento de dados; padrões de metadados estabelecem somente regras de uso para os dados.
Dados	, , 92
Metadados	<SQUARE> <SQUARE> <AREA>92<AREA>

Medida	Intemporal
Descrição	Disponibilidade dos dados com atualizações precisas sem atrasos na apresentação dos mesmos.
Justificativa	A disponibilidade do dado não pode ser mensurada por meio do padrão de metadados; recursos como esses podem ser analisados em ferramentas utilizadas para gerenciar o padrão em uma linguagem de especificação.
Dados	Não aplicado.
Metadados	Não aplicado.
Medida	Relevância
Descrição	Medida relacionada com a confiança do dado representado em uma determinada instância, especificamente, em dados arrolados com a significância no contexto onde o mesmo é gerado e utilizado.
Justificativa	A utilidade dos dados é definida por meio do mapeamento das classes de dados, de acordo com os processos nos quais foram derivados; o padrão de metadados proporciona somente uma interface para os dados.
Dados	T-1, 92
Metadados	<SQUARE>T-1<SQUARE> <AREA>92<AREA>
Medida	Segurança de acesso
Descrição	Disponibilização de mecanismos para controle de acesso a diferentes tipos de usuários.
Justificativa	Metadados são padrões e não disponibilizam recursos para ocultar, criptografar ou restringir determinado tipo de acesso; nesse sentido, é necessário implementar esse recurso na linguagem utilizada para a sua formalização.
Dados	-16.083754
Metadados	<LATITUDE>-16.083754</LATITUDE>

Outro ponto importante referente a intercâmbio de dados pode ser observado em aspectos relacionados com desempenho e espaço de armazenamento. Nesse sentido, foram utilizadas duas métricas para mensurar tempo de resposta (*Response to Request em ms*) e tamanho dos dados (*Data Size em Mb*), possibilitando a verificação do tempo demandado para atender às requisições e o espaço físico necessário ao arquivamento das estruturas. No Gráfico 12 é ilustrado o resultado da métrica usada para quantificar o tempo de resposta para cada implementação e o tamanho físico demandado para armazenar a base de dados consumida na implementação dos dois paradigmas. O tempo de resposta mostrou mais eficiência na solução com metadados, com taxa de 42% mais eficiente que a implementação

sem metadados, justificado pela simplicidade do protocolo utilizado para implementar a interface com o padrão de metadados. Apesar da burocracia imposta pelo padrão de metadados a infraestrutura resume-se a um conteúdo textual sem dados binários, facilitando a resposta de implementações quanto a processamento e tráfego de rede.

Gráfico 12 - Tempo de resposta (ms).

O espaço físico em unidades de armazenamento de dados secundário foi similar nas duas implementações, com demanda maior para a estrutura de dados com metadados, incrementando, em aproximadamente 10%, em razão da infraestrutura de *tags* exigida pelo padrão, conforme ilustrado no Gráfico 13.

Gráfico 13 - Espaço de armazenamento (Mb).

7.2.5.3 Métricas de Qualidade de Dados

A terceira etapa da análise dos resultados refere-se às análises entre os três padrões de dados e os testes foram realizados por meio de comparações entre os padrões de trocas de dados CSV, dicionários de dados e o padrão de metadados proposto. Os dados apresentados na primeira coluna do Quadro 5 ilustram as deficiências quanto à legibilidade em formatos sem identificações ou referências dos dados armazenados. Apesar de os sistemas desenvolvidos para o segmento agrícola, em geral, disponibilizarem recursos para exportar bases de dados completas ou em fragmentos, conforme exposto nos Gráficos 2 e 3, os formatos disponíveis para exportação são proprietários. O formato CSV exibido na primeira coluna do Quadro 11 é comum para esse tipo de intercâmbio de dados; sua estrutura simples e independente é composta por colunas separadas por vírgula ou ponto-e-vírgula em um arquivo padrão *Unicode*. A forma exposta na segunda coluna utiliza definições de dados disponíveis em dicionários específicos em sistemas gerenciadores de bancos de dados, para identificar as características dos dados a serem representadas e a melhora, quanto à legibilidade na apresentação dos dados, é significativa e satisfatória. Entretanto, a visualização dos dados é restrita a ferramentas específicas com suporte para a interpretação dos metadados do dicionário de dados.

Quadro 25 - Comparativo entre padrão csv e dicionário de dados.

Formato CSV	Dicionário de Dados					
	FARMNAME	IDSQUARE	IDFARM	SQUARENAME	SQUAREAREA	SQUARELAT
Colorado,1,1,T-1,92,-16,083754,-54,930124	Colorado	1	1	T-1	92	-16,083754
Colorado,2,1,T-2A,110,-16,069568,-54,928064	Colorado	2	1	T-2A	110	-16,069568
Colorado,3,1,T-2B,110,-16,074187,-54,913473	Colorado	3	1	T-2B	110	-16,074187
Colorado,4,1,T-3A,120,-16,066104,-54,906263	Colorado	4	1	T-3A	120	-16,066104
Colorado,5,1,T-3B,45,-16,06363,-54,895449	Colorado	5	1	T-3B	45	-16,06363
Colorado,6,1,T-4A,85,-16,098928,-54,919653	Colorado	6	1	T-4A	85	-16,098928
Colorado,7,1,T-4B,114,-16,091341,-54,913988	Colorado	7	1	T-4B	114	-16,091341
Colorado,8,1,T-5A,113,-16,081774,-54,90695	Colorado	8	1	T-5A	113	-16,081774
Colorado,9,1,T-5B,115,-16,073197,-54,898195	Colorado	9	1	T-5B	115	-16,073197
Colorado,10,1,T-6,111,-16,063795,-54,886007	Colorado	10	1	T-6	111	-16,063795
Colorado,11,1,T-7A,131,-16,111132,-54,921885	Colorado	11	1	T-7A	131	-16,111132
Colorado,12,1,T-7B,117,-16,103051,-54,911757	Colorado	12	1	T-7B	117	-16,103051
Colorado,13,1,T-8A,116,-16,095464,-54,902315	Colorado	13	1	T-8A	116	-16,095464
Colorado,14,1,T-8B,65,-16,085403,-54,890471	Colorado	14	1	T-8B	65	-16,085403
Colorado,15,1,T-9A,86,-16,139661,-54,918966	Colorado	15	1	T-9A	86	-16,139661
Colorado,16,1,T-9B,115,-16,128943,-54,903688	Colorado	16	1	T-9B	115	-16,128943
Colorado,17,1,T-9C,55,-16,150709,-54,887381	Colorado	17	1	T-9C	55	-16,150709
Colorado,18,1,T-10A,102,-16,119543,-54,891844						
Colorado,19,1,T-10B,115,-16,110967,-54,881029						
Colorado,20,1,T-10C,68,-16,096289,-54,874506						

	Colorado	18	1	T-10A	102	-16,119543	-54,891844
	Colorado	19	1	T-10B	115	-16,110967	-54,881029
	Colorado	20	1	T-10C	68	-16,096289	-54,874506

O padrão resultante da formalização da ontologia é ilustrado na Figura 39, onde o padrão de metadados é utilizado para estruturar dados referentes a quadras (*square*) em uma fazenda (*farm*); o fragmento formaliza dados como nome, área em hectares, latitude e longitude para fornecer a localização ilustrativa de cada quadra. Os dados são os mesmos presentes na Figura 39.

Figura 39 - Dados formalizados utilizando o padrão de metadados com XML.

De forma complementar, o padrão de metadados contribui com a melhoria no processo da qualidade de dados, sendo que o uso de um padrão único em todos os processos possibilita customizações no desenvolvimento de interfaces para a gestão de dados. A avaliação de aspectos relacionados com qualidade de dados é realizada por meio de dimensões estabelecidas em referências como McGilvray

(2008). Dimensões de qualidade de dados são aspectos ou características para mensurar dados e informações. As consequências de dados com pouca qualidade são observadas no cotidiano, sem a identificação de suas causas. Já a qualidade dos dados também tem consequências diretas na geração de informações úteis para o gerenciamento organizacional.

A infraestrutura de *tags* disponíveis no padrão de metadados possibilita a formatação dos dados em uma estrutura consistente e simples, contribuindo com a visibilidade e inteligibilidade por todos os envolvidos no uso dos dados. A Figura 40 resume os três tipos de padrões para exposição de dados discutidos na pesquisa: uma estrutura baseada em CSV, sem formatação específica, um formato proprietário em banco de dados e, finalmente, os dados são apresentados, utilizando o padrão de metadados proposto. Apesar de os dados estarem organizados com infraestrutura de dicionário de dados, dispostos de forma tabular, em colunas com relativa inteligibilidade, seu compartilhamento é restrito a ferramentas gerenciadoras de dados e seus complementos.

Figura 40 - Integridade de dados em fragmento de metadados.

A clareza da estrutura é importante para a identificação do significado dos dados mesmo que estejam incompletos, e a *tag FarmName* descreve, previamente, o significado dos dados representados nesse contexto. Dessa forma, um dado

incompleto como “COLORADO” pode ser diagnosticado como “FAZENDA COLORADO”, pois a característica autodescritiva do padrão de metadados possibilita essa identificação. Outra contribuição significativa do padrão de metadados está no conhecimento prévio da hierarquia dos dados representados. Isso porque entender a estrutura contribui com a recuperação de dados seguindo caminhos hierárquicos e possibilita a criação de aplicações para geração de informações.

7.3 DISCUSSÃO DOS RESULTADOS

As análises dos resultados dividem-se em três partes: desenvolvimento de *software*, intercâmbio de dados e qualidade de dados.

Os resultados das análises referentes ao desenvolvimento de *software* foram apresentados nos Gráficos 5 a 10. O primeiro efeito exposto refere-se à complexidade ciclomática, com foco na quantidade de estruturas condicionais utilizadas nas implementações. A implementação com emprego de metadados mostrou redução nessa métrica, em média 30% menor, em relação à estrutura sem o uso de metadados. A justificativa está na ausência de recursos relacionados com interface para execução do protótipo com metadados, pois, nesse tipo de prática, não há interface para acesso de dados, cujos serviços são desenvolvidos posteriormente, de forma customizada. Dessa forma, não são utilizados códigos para construção de estruturas de interfaces, resultando, assim, em redução da complexidade da implementação.

A métrica mostrada no Gráfico 6 avalia a quantidade de linhas de código usada em cada implementação. A ausência de interface entre as aplicações com características de serviços, ou implementação distribuída nesse contexto com uso de metadados, proporcionou redução de aproximadamente 70% em relação à realização sem metadados, na totalidade de linhas de código.

A quantidade de membros criados em classes pode ser observada no Gráfico 7, com redução de aproximadamente 77% no protótipo utilizando metadados.

Os códigos com metadados foram desenvolvidos para atender, estruturalmente, somente a solicitações sem o uso de interface com o usuário; dessa forma, as classes criadas têm a função de receber requisições e devolver respostas sem componentes gráficos ou de interface.

Complementarmente ao tamanho do código apresentado no Gráfico 7, o Gráfico 8 expõe a quantidade média de caracteres por linha. A implementação com metadados apresentou significativa redução de 10% na quantidade de código fonte por linha e também na quantidade de códigos para a construção dos protótipos, demonstrando melhora tanto no processo de desenvolvimento quanto em futuras manutenções.

Funções representam um recurso para o desenvolvimento de *softwares* computacionais. Apesar de não existirem métricas específicas para estabelecer a quantidade proporcional adequada para o uso de funções em códigos fontes, é recomendável a utilização sempre que possível tanto isoladamente quanto vinculada a classes. Nesse segundo caso, refere-se aos recursos usados em programação orientada a objetos. Os valores apontados no Gráfico 9, relacionados à quantidade de parâmetros para entrada de valores e retorno por referência, foram os mesmos nos códigos para os dois tipos de protótipos, tanto com metadados quanto sem metadados, não apresentando melhora ou perda nesse quesito. Além da implementação de funções discutida, outro ponto relevante diz respeito ao uso e à quantidade total de funções implementadas. A prática de metadados exibiu valores superiores quanto ao número total de parâmetros, com aproximadamente 50% a mais em relação à implementação sem metadados, e isso corresponde a uma situação pontual, não refletindo, de forma genérica, para as demais práticas. Os valores referentes à quantidade total de funções instaladas apresentou redução de 25% com a implementação com metadados, conforme Gráfico 10. Isso se deve ao fato de o código constituir menor complexidade, justificado pela ausência de interface final para manipulação.

A verificação do comportamento do intercâmbio de dados nos protótipos desenvolvidos apresenta-se em dois resultados: o primeiro referente à análise qualitativa do padrão de metadados, frente às métricas de qualidade de *software* e o

segundo, relacionado às métricas de armazenamento e resposta. A análise qualitativa foi realizada utilizando um conjunto reduzido de dados para verificar a adequação do padrão junto às diferentes características necessárias para o seu uso. Nesse contexto, o uso do padrão atendeu às dimensões: acessibilidade, completude, conformidade, especificação de dados, facilidade de uso e manutenibilidade, fonte, objetividade, qualidade de apresentação, transatabilidade e validade. As dimensões usadas para a análise foram selecionadas por tratarem das características de apresentação ou armazenamento. Não foram consideradas as dimensões: acurácia, cobertura de dados, consistência, duplicações, empobrecimento de dados, integridade, intemporal, relevância e segurança de acesso. Essas dimensões não são favoráveis ao uso do padrão e representam regras de integridade, precisão de dados e segurança. Nesse sentido, o padrão não traz domínios rígidos para a imposição de regras que contemplem essas dimensões. Tais regras são, usualmente, impostas por meio de motores de banco de dados ou linguagens de programação.

A segunda parte dos resultados, referente ao intercâmbio de dados, pode ser observada por meio de duas métricas: tempo de resposta e espaço de armazenamento. O tempo de resposta medido em milissegundo, apesar de ter a carga burocrática do padrão de metadados embutido na apresentação dos dados, obteve resultado melhor. Os testes realizados foram implementados em bancos de dados com suporte a metadados, formalizados em *XML*; dessa forma, a ausência de camadas de aplicações intermediárias para realizar o tratamento dos dados retornados, e posterior apresentação, não é necessária, resultando em ganho de tempo para o solicitante, que pode ser um *shell* em banco de dados ou uma ferramenta de consulta mais complexa, conforme o Gráfico 12. Os dados no Gráfico 13 refletem uma característica comum em dados formatados em padrões de metadados que incorporam as *tags*. Com isso, o espaço total necessário para o armazenamento aumenta significativamente.

A terceira parte da discussão dos resultados verifica o comportamento dos dados em três padrões para formalização de dados: CSV, dicionários de dados e o padrão de metadados proposto. A análise é realizada sobre o comportamento

estrutural de um conjunto de dados referentes a subáreas de cultivo em uma fazenda. As colunas de dados representam: nome da fazenda, identificação da subárea de cultivo, identificação da fazenda, nome da subárea, área em hectares da subárea, latitude e longitude de cada subárea. Os dois últimos dados descrevem somente um ponto geográfico. O padrão CSV estrutura os dados apenas em colunas separados por vírgula (,) ou ponto e vírgula (;), em uma estrutura simples e compatível com diversas plataformas operacionais e ferramentas de gerenciamento de dados. A limitação quanto a esse padrão está na ausência de metadados para estabelecer, singularmente, o significado dos dados. Assim, o seu uso não é intuitivo e não atende a dimensões como completude e especificação de dados, entre outras. O segundo padrão avaliado considera o dicionário de dados disponível em sistemas gerenciadores de banco de dados. Nesse caso, o padrão disponibilizado pelo próprio banco de dados mostra-se eficiente quanto à recuperação e apresentação de dados; entretanto, o seu uso está restrito ao próprio banco de dados ou a ferramentas com recursos legados ou específicos.

Assim, a extensibilidade de uso fica comprometida ou limitada. Por fim, o padrão de metadados proposto para a cotonicultura apresenta *tags* em uma estrutura *XML* simples, com conteúdo armazenado em arquivos físicos com padrão *Unicode*; por isso, o uso não se limita a uma plataforma ou ferramenta específica. A interpretação do conteúdo pode ser realizada sem o auxílio de ferramentas computacionais específicas, mostrando maior flexibilidade de uso.

Os Gráficos 14, 15 e 16 revelam um resumo da análise dos resultados nas três etapas supracitadas. O Gráfico 14 quantifica as métricas apresentadas para o desenvolvimento de *software* (complexidade ciclomática, quantidade de linhas de código, número de membros, quantidade média de caracteres por linha, quantidade média de parâmetros de entrada, quantidade média de parâmetros de saída, parâmetros por função e número total de funções), simplesmente somando o total de itens beneficiados com o padrão de metadados e não beneficiados. Dessa forma, foram quantificadas oito métricas das quais somente duas obtiveram resultados melhores ou iguais aos apresentados na implementação com metadados. O propósito é fornecer uma visão objetiva a respeito da contribuição do padrão.

Gráfico 14 - Resumo de métricas de desenvolvimento de *software*.

Os valores apresentados no Gráfico 15 resumem, quantitativamente, os resultados da análise referente ao uso das dimensões de intercâmbio de dados atendidas com o padrão de metadados proposto. A análise quantitativa partiu da verificação da adequação do padrão de metadados para as 19 dimensões discutidas no capítulo 4.2.5.2, cujo padrão proposto atendeu apenas a 10 dimensões, mostrando, quantitativamente, a qualidade estrutural do mesmo.

Gráfico 15 - Resumo de métricas de intercâmbio de dados.

O Gráfico 16 quantifica as dimensões de qualidade de dados conforme discussão da terceira parte da análise dos resultados no item 4.2.5.3, verificando aspectos como: nome de colunas, domínio de dados, necessidade de ferramentas

específicas e também a possibilidade de uso em diversas plataformas operacionais. Esse Gráfico apresenta a quantidade de itens contemplados por padrão, sendo 1, 3 e 4 ou 25% para cada padrão respectivamente. Com isso, a adequação do padrão de metadados aos itens analisados percentualmente resultou em: 25% para o padrão CSV, 75% para o dicionário de dados e 100% para o padrão de metadados proposto.

Gráfico 16 - Resumo de métricas de qualidade de dados.

O objetivo da prova de conceito é avaliar a evolução da qualidade dos dados no segmento agrícola abordado. A pesquisa enfatiza os problemas no intercâmbio de dados, propondo um padrão de metadados para contribuir com a redução dessa deficiência. Melhorias na qualidade de dados resultantes do aperfeiçoamento do intercâmbio de dados são benefícios complementares, observados por meio de métricas de qualidade de dados.

8. CONCLUSÃO

A formalização do conhecimento por meio da ontologia foi resultado de uma grande coleta de dados tanto bibliográficos quanto presenciais, diversas dificuldades relacionadas com o acesso a pessoal técnico especializado foram encontradas tanto no processo de coleta quanto na validação de fragmentos do modelo proposto. É característica no segmento agrícola a dificuldade em obtenção de dados de forma unificada, em geral há uma abordagem vertical quanto a cada processo dificultando assim a formalização de modelos mais abrangentes e completos.

O conhecimento representado na ontologia forneceu um grau de inteligibilidade suficiente para mapeamento dos processos da cotonicultura para a construção do padrão de metadados. O padrão atende às especificações de dados utilizados na cotonicultura, nos processos entre o preparo do solo e o beneficiamento, conforme averiguado na prova de conceito e complementarmente por meio de validações presenciais junto a técnicos de áreas específicas relacionadas a agricultura. A contribuição conceitual da pesquisa está na construção da ontologia formalizada por meio de um padrão de metadados, disponibilizando recursos para o desenvolvimento de interfaces em soluções computacionais.

As análises realizadas por meio da prova de conceito possibilitaram a avaliação do padrão de metadados, do ponto de vista de implementação e o uso de protótipos para realização dos testes mostrou a viabilidade estrutural do padrão na formalização de dados, resultando na melhoria da qualidade dos mesmos. Outro ponto importante, relacionado com efetivações por meio do padrão proposto, está na criação de interfaces para aplicações computacionais, uma vez que o padrão contribui com a padronização de respostas a requisições de dados, favorecendo o uso ou a publicação de serviços.

O uso efetivo do padrão de metadados proposto como contribuição para a melhoria do intercâmbio de dados depende em grande parte da aceitação comercial do mesmo junto a desenvolvedores de soluções computacionais. Para isso validações e homologações do padrão por grupos e associações relacionados com a produção da fibra do algodão são essenciais. Iniciativas nesse sentido estão sendo realizadas para a consolidação efetiva do padrão proposto.

O aporte da pesquisa para agricultura está no uso do padrão para homogeneização e intercâmbio de dados, seja de forma completa ou fragmentada (conjunto de tabelas), destacando que essa homogenia é a base de diversas aplicações para mineração de dados e implementação de soluções para auxílio na tomada de decisões gerenciais, desde o plantio até a comercialização na cotonicultura. As soluções computacionais mencionadas são beneficiadas por meio do novo padrão de metadados, formalizado a partir da ontologia proposta. Nesse contexto, a pesquisa analisou a contribuição para melhoria do intercâmbio e qualidade de dados com um conjunto de dados agrícolas reais, verificando o apoio do padrão de metadados. Conforme discutido no capítulo 5, os resultados apresentaram dados positivos e satisfatórios quanto à viabilidade de uso do padrão proposto nas análises referentes ao: desenvolvimento de software, intercâmbio de dados e qualidade de dados.

8.1 Trabalhos Futuros

A ontologia formalizada na pesquisa contribuirá com o desenvolvimento de novos trabalhos relacionados com qualidade e intercâmbio de dados, viabilizando também a criação ferramentas computacionais (*software*) e publicações em revistas técnicas específicas.

1) Mapeamento dos subprodutos do algodão por meio de ontologia para homogeneização de dados: a pesquisa abordou a representação do conhecimento referente aos processos para um subproduto do algodão, nesse contexto a fibra. Diversos outros subprodutos também derivam da cotonicultura os mesmos problemas relacionados a intercâmbio estão presentes. O uso de ontologia aplicada para formalização de outros subprodutos pode representar uma importante contribuição acadêmica.

2) Ontologia Aplicada a Construção de Interfaces em *Middlewares* para Sincronismo de Bancos de Dados na Agricultura: uma característica importante em soluções computacionais para o agronegócio está na heterogeneidade de sistema locais e principalmente remotos, grande parte do problema está relacionado com o

empobrecimento de dados provenientes das dificuldades de sincronização entre as bases de dados distribuídas em locais geograficamente distantes. O uso de ontologias para a formalização de requisitos entre esses sistemas e posteriormente implementação de middlewares para melhorar esse tipo de intercâmbio de dados pode também representar uma importante contribuição.

9. REFERÊNCIAS

AGXML GROUP. agXML Overview. 2010. Disponível em: <<http://www.agxml.org/Default.aspx?tabid=293>>. Acesso em: nov. 2011.

AMPASUL. Manual de Orientação Técnica para o Algodão em Mato Grosso do Sul. Associação Sul Mato-Grossense dos Produtores de Algodão. 2011. Disponível em: <<http://www.ampasul.com.br>>. Acesso em: ago. 2011.

BARTOL, T. Assessment of classification and indexing of an agricultural journal based on metadata in AGRIS and CAB Abstracts databases. International Journal of Metadata, Semantics and Ontologies, In: International Journal of Metadata, Semantics and Ontologies, 2009. Geneva. **Proceedings...** Geneva: 2009. p. 2009.

BITTNER, T.; DONNELLY M. **Ontology and Semantic Interoperability**. Institute for Formal Ontology and Medical Information Science. Saarbrücken/DE. 2004.

BORTOLINI, C. **Preparo de Perfil de Solo com Equipamentos Mecânicos Visando Cultivo do Algodão Safrinha**. Lucas do Rio Verde/MT. Fundação de Apoio a Pesquisa e Desenvolvimento Integrado Rio Verde. 2005.

BUAINAIN, A. Cadeia Produtiva do Algodão. **Série Agronegócios**, v4, p 47-80, 2007.

CASTRO, A. **Análise da Competitividade de Cadeias Produtivas**. 2000. Disponível em: <<https://www.suframa.gov.br/download/publicacoes/accp.pdf>>. Acesso em: nov. 2009.

CONAB. Mapa da Produção de Algodão no Brasil. 2010. Disponível em: <<http://www.conab.gov.br>>. Acesso em: jul. 2011.

COMMUNITIES, R. Dublin Core Metadata Element Set. 2010. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: abr. 2010.

CORRÊA, S.; COUTO, E. **A História do Algodão no Brasil e seu Desenvolvimento no Estado do Mato Grosso, o Atual Maior Produtor do País, Instituto de Economia.** Instituto de Economia. Universidade Federal de Uberlândia. Uberlândia/MG. 2003.

COUGHLIN, T. M.; LINFOOT, S. L. A novel taxonomy for consumer metadata. In: International Conference on Consumer Electronics, 2010, Las Vegas. **Proceedings...** Las Vegas: 2010. p. 1-3.

CRUZ, I., et al. An Ontology-based Framework for XML Semantic Integration. 2004. Disponível em: <http://www.cs.uic.edu/~advis/publications/dataint/ideas04.pdf>. Acesso mai. 2013.

DATA INTERCHANGE, Electronic Data Interchange. 2005. Disponível em: <http://www.datainterchange.com/downloads/brochures/whitepaper-what-is-edi.pdf>>. Acesso em: jul. 2013.

DENNY, M., 2013. **Table from Ontology Editor Survey.** Marine Metadata Initiative. 2004. Disponível em: http://www.xml.com/2004/07/14/examples/Ontology_Editor_Survey_2004_Table_-_Michael_Denny.pdf>. Acesso em: jan. 2013.

FERGUSON, R., et al. Soil Sampling for Precision Agriculture. **Precision Agriculture.** p. 2-4. 2009.

FERNANDES, A. The social and environmental impacts of industrial agriculture in the Legal Amazon. In: XIV Simpósio Brasileiro de Sensoriamento Remoto, 2009, Natal. **Anais...** Natal: 2009. p. 3-5.

FILETO, R., et al. **Using Domain Ontologies to Help Track Data Provenance.** Campinas/SP: Embrapa Tecnologia. 2011.

GUARINO, N. Formal Ontology and Information Systems. In: Formal Ontology and Information Systems, 1998, Amsterdam. **Proceedings...** Amsterdam: 1998. p. 3-15.

IKEDA, M., et al. **Task ontology: Ontology for building conceptual problem solving models.** Hyogo/JP. 1999.

INAN, H. Data model for the collaboration between land administration systems and agricultural land parcel identification systems. **Journal of Environmental Management.** v. 1, p. 1-3, 2010.

INITIATIVE, D. Dublin Core Metadata Element Set. 2010. Disponível em: <<http://dublincore.org/>>. Acesso em: jul. 2010.

JONES, C., A Short History of the Lines of Code (LOC) Metric. 2008. Disponível em: <www.gilb.com/dl187>. Acesso em: mai. 2013.

Li, M. et al. Learning Ontology from Relational Database. In: Fourth International Conference on Machine Learning and Cybernetics, 2005, Guangzhou. **Proceedings...** Guangzhou: 2005. p. 1-3.

MALONE, J.; Parkinson, H. Reference and Application Ontologies. 2010. Disponível em: <<http://ontogenesis.knowledgeblog.org/295>>. Acesso em: jan. 2014.

MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Agronegócio Brasileiro: Uma Oportunidade de Investimento. 2010. Disponível em: <<http://www.agricultura.gov.br>>. Acesso em: mar. 2010.

MARTINI, D. et al. A service architecture for facilitated metadata annotation and resource linkage using agroXML and restful web services. **Metadata and Semantic Research Communications in Computer and Information Science**, v 46, p 257-262, 2009.

MCCABE, T. A Complexity Measure. **IEEE Transactions on Software Engeneering**. v. 2, n. 4, p. 1-3, Dez. 1976.

MCGILVRAY, D. **Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information**. MA, USA: Morgan, 2008.

MUSEN, M. Domain Ontologies in Software Engineering: Use of Protégé with the EON Architecture. 1995. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.6458&rep=rep1&type=pdf> . Acesso em: jan.2014.

NARA, National Archives and Records Administration E-Gov Electronic Records Management. Initiative Recommended Practice: Developing and Implementing an Enterprise-wide Electronic Records Management (ERM) Proof of Concept Pilot. 2006. Disponível em: <http://www.archives.gov/records-mgmt/policy/pilot-guidance.html>>. Acesso em: out. 2013.

NATLACEN, M.; AL, E. The AGRIS Application Profile for the International Information System on Agricultural Sciences and Technology Guidelines on Best Practices for Information Object Description. 1998. Disponível em: <http://www.fao.org/docrep/008/ae909e/ae909e00.htm>>. Acesso em: set. 2010.

NISO. National Information Standards Organization. **Understanding Metadata**. 2004. Disponível em: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: fev. 2011.

O'CALLAGHAN, R. Electronic Data Interchange Concepts and Issues. 2005. Disponível em: <http://archive.nyu.edu/bitstream/2451/14222/1/IS-95-15.pdf>>. Acesso em: ago. 2013.

PRESSMAN, Roger S. **Software engineering: a practitioner's approach**. NY, USA: McGraw-Hill, 2009.

REICHERT, L. **Gestão de Propriedades Familiares Rurais**. Embrapa do Centro de Pesquisa Agropecuário de Clima Temperado de Pelotas. Universidade Católica de Pelotas. 2001. Dissertação (Mestrado).

RICHETTI, A. et al. **Sistemas de Produção**. 2003. Disponível em: <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Algodao/AlgodaoCerrado/index.htm>>. Acesso em: set. 2011.

RILEY, C. Quality Measurements: Improved High Volume Instrument Elongation Measurements. 1997. Disponível em: <<http://journal.cotton.org>>. Acesso em: fev. 2010.

SANCHEZ-ALONSO, S.; SICILIA, M. A. **Using an AGROVOC-based ontology for the description of learning resources on organic agriculture**. Metadata and Semantics. 2009.

SANTANA, A, Software de Gestão, Estudo de Mercado SEBRAE/ESPM. 2008. Disponível em: < [http://bis.sebrae.com.br/GestorRepositorio/ARQUIVOS_CHRONUS/bds/bds.nsf/B316EA7311BA4E448325753E005FA07D/\\$File/NT0003DB22.pdf](http://bis.sebrae.com.br/GestorRepositorio/ARQUIVOS_CHRONUS/bds/bds.nsf/B316EA7311BA4E448325753E005FA07D/$File/NT0003DB22.pdf)>. Acesso em: set. 2013.

SCHMITZ, M. et al. agroXML: Enabling Standardized, Platform-Independent Internet Data Exchange in Farm Management Information Systems. **Metadata and Semantic Research Communications in Computer and Information Science**, p. 463-468, 2009.

SOFIATTI, V. et al. **Algodão em Pluma**. Campina Grande: Embrapa Algodão. 2009.

SOMMERVILLE, I. **Software Engineering**. NJ, USA:Pearson, 2009.

TDWG, TDWG Standards. 2007. Disponível em: <<http://www.tdwg.org>>. Acesso em: mar. 2010.

URS, S. Ontology-based Knowledge Organization Systems in Digital Libraries: A Comparison of Experiments in OWL and KAON Ontologies. 2007. Disponível em: <http://www.academia.edu/695075/Ontology-based_Knowledge_Organization_Systems_in_Digital_Libraries_A_Comparison_of_Experiments_in_OWL_and_KAON_Ontologies>. Acesso em: set. 2013.

VALKENGOED, E. GEO-SPATIAL DATA ISSUES. Europe Union. Galileo Joint Undertaking. 2007. Disponível em: <http://due.esrin.esa.int/prjs/Results/131-176-149-30_2008229113128.pdf>. Acesso em: ago. 2012.

WANG, H. et al. Dublin Core-based metadata extracting algorithm from RDF-based information elements. 2008. Disponível em: <<http://dx.doi.org/10.1109/WKDD.2008.113>>. Acesso em: ago. 2010.

WANG, R., et al. **Data Quality**. MA, USA: Kluwer, 2002.

WEICK, C. Agribusiness technology in 2010: directions and challenges. **Technology in Society**, v. 23, p. 59-72, 2001.

ZONG-YONG, L. et al. The Domain Ontology and Domain Rules Based Requirements Model Checking. **International Journal of Software Engineering and Its Applications**, v. 1, p. 1-3, 2007.

YAMAGUCHI, T. Constructing Domain Ontologies Based on Concept Drift Analysis, 1999. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.3113&rep=rep1&type=pdf>>. Acesso em: dez. 2013.