

Interoperabilidade: um estudo sobre padrões de metadados XML aplicados no domínio agrícola com potencial para uso na indústria canavieira**Fábio de Paula Santos¹**
Edson Murakami²

Resumo: Apesar da adoção de ERP's (*Enterprise Resource Planning*) especializados em cana-de-açúcar ter crescido, ainda existem muitos sistemas legados e arquivos que utilizam dados não estruturados. Com isto, temos um aumento da diversificação dos sistemas de informação utilizados na gestão da produção canavieira, levando a problemas de integração devido à heterogeneidade das fontes e formatos dos dados. Também é crescente a necessidade dos envolvidos na cadeia produtiva canavieira de acessarem essas informações, o que torna imprescindível a padronização dos dados a serem trocados entre eles. Neste artigo são apresentadas algumas iniciativas de padronização nos formatos dos dados que têm sido conduzidas por pesquisadores e profissionais do domínio agrícola e que têm potencial de serem utilizados no domínio da cana-de-açúcar. Destacamos a AgXML, a AgrisAP, a AgroXML e a PAML. A última é uma iniciativa brasileira de padronização de dados para o domínio da Agricultura de Precisão. Os documentos gerados pela PAML são usados como padrão para troca de dados entre aplicações para agricultura de precisão, o que possibilita que aplicações heterogêneas se comuniquem aproveitando a infraestrutura da *Internet* e reduzindo o problema de integração de sistemas e fontes de dados, problema comum também nos sistemas para gestão da produção canavieira.

Palavras-chave: Agronegócio. Interoperabilidade. XML.

Abstract: Despite the growth of the adoption of ERP's (*Enterprise Resource Planning*) specialized in sugar cane, there are still many legacy systems and files that use non-structured data. Thus, we have an increase of the diversification of the information systems used in the sugarcane production management, heading to integration problems due to the heterogeneity of the data sources and sizes. The need of the involved in the sugarcane productive chain to access these informations are also increasing, which makes necessary the standardization of the data that will be changed among them. This paper present some standardization initiatives in the format of the data that have been conducted by researches and professionals if the agricultural domain with potential to sugarcane domain. We highlight AgXML, AgrisAP, AgroXML and PAML. The last is a Brazilian initiative of data standardization for the Precision Agriculture domain. The generated documents by the PAML are used as a standard for the data interchange between applications, which makes possible that the heterogeneous applications

¹ FATEC Itu – fabio@pssoftware.com.br (autor para correspondência)

² IFECT de São Paulo, *Campus* Salto – murakami@ifsp.edu.br

communicate using the Internet infrastructure and reduce the integration problem in many systems and data source, a common problem also in systems for sugarcane production management.

Keywords: Agribusiness. Interoperability. XML.

1 Introdução

Segundo Abrahão (2009), nos últimos 10 anos notou-se um aumento na utilização de Sistemas de Informação nas operações agrícolas canavieiras, com consequente aumento da coleta, processamento e uso dos dados. No entanto, apesar da adoção de ERP's (*Enterprise Resource Planning*) especializados em cana-de-açúcar ter crescido, ainda existem muitos sistemas legados que utilizam, por exemplo, planilhas eletrônicas, o que aumenta a diversificação dos sistemas de informação utilizados na gestão da produção canavieira.

A diversidade de sistemas aumenta os problemas de integração, devido à heterogeneidade das fontes e formatos dos dados. Além disso, é crescente a necessidade, para os envolvidos na cadeia produtiva canavieira, de acesso às informações desses sistemas, o que torna imprescindível a padronização dos dados a serem trocados entre eles.

Este artigo destaca várias iniciativas de padronização dos formatos dos dados, que de alguma forma adicionam semântica para facilitar a troca de informações entre sistemas, e que têm sido conduzidas por pesquisadores e profissionais do domínio agrícola. Por exemplo, a AgXML (AgXML, 2004) para troca genérica de dados na área agrícola, a AgrisAP (AgrisAP, 2005) para o domínio da soja e grãos em geral, a PAML (*Precision Agriculture Markup Language*) (MURAKAMI et al., 2007) para o domínio da agricultura de precisão, além do AgroXML (DOLUSCHITZ, et al., 2005) para o domínio genérico da agricultura, o qual tem evoluído bastante e poderia ser utilizado para qualquer domínio agrícola.

Este artigo sugere soluções de integração de aplicações no domínio agrícola que usam a linguagem de marcação XML (*eXtends Markup Language*), para troca de informações entre sistemas. Tais soluções têm grande potencial e podem contribuir na interoperabilidade dos sistemas envolvidos no domínio da cana-de-açúcar.

2 Padrões

Os computadores eletrônicos surgiram a partir de aplicações militares para quebra de códigos cifrados em mensagens (decriptografia), nas décadas de 30 e 40 do século passado. Após a segunda guerra mundial esta tecnologia passou para o mundo dos negócios. Com isso, cada empresa fornecedora de soluções, elaborava seu próprio padrão para a troca de informações e formatação dos dados. Um exemplo desta troca seria enviar junto com o texto a ser transmitido uma marcação “<Tamanho 12>” e, o software responsável pela leitura, entenderia que o conteúdo entre os símbolos “<” e “>” não deveria ser exibido e sim, iniciar o procedimento para a impressão no tamanho de fonte 12. Notou-se então que, para uma troca de dados bem sucedida, os lados receptores e emissores devem seguir um mesmo padrão. O padrão GML (*Generalized Markup Language*), introduzido na década de 60 pela IBM é um exemplo desses padrões.

Na Figura 1 são mostrados alguns padrões da indústria surgidos ao longo do tempo.

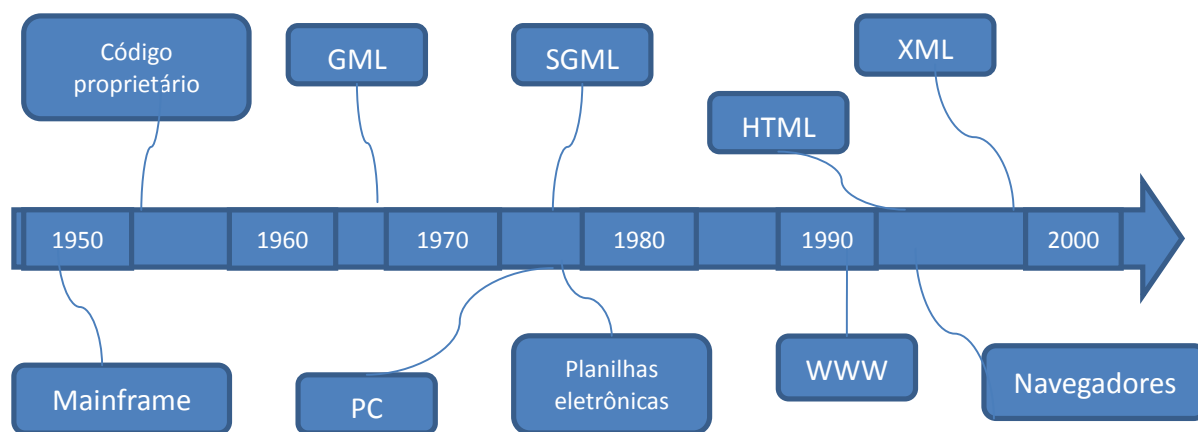


Figura 1- Padrões da Indústria. Adaptados pelos autores de (BERGERON, 2003).

Com a popularização do uso dos computadores pessoais (PC's) em meados da década de 80, ficou evidente a necessidade da criação de padrões para a troca de documentos.

Na década de 90 surge a HTML (*HiperText Markup Language*), que permitiu a criação de páginas para serem disponibilizadas na rede mundial de computadores. No final dessa década, em 1998, a partir da necessidade de páginas dinâmicas na *internet*, surgiu a XML (*eXtend Markup Language*), que permitia o acesso a banco de dados e a troca de dados entre aplicações. A partir da XML diversas derivações para domínios específicos surgiram para facilitar a busca e recuperação de documentos na *Internet* e principalmente para troca de informações entre aplicações heterogêneas inerentes à natureza da *Internet*.

Portanto, a XML tem papel importante na interoperabilidade entre aplicações de qualquer domínio de negócio, particularmente na cadeia produtiva que envolve vários interessados. As experiências de padronização, com XML, em diversos domínios da agricultura têm mostrado que é possível que aplicações heterogêneas convivam entre si.

Porém, para que todas as aplicações envolvidas numa cadeia produtiva possam “falar a mesma língua” é necessário um vocabulário comum e padronizado com capacidade de extensão, como definido nos padrões citados acima, AgXML, AgrisAP, PAML e AgroXML. Todos esses padrões são baseados na XML, o que comprova que a cadeia produtiva da cana-de-açúcar poderia se beneficiar dessa interoperabilidade se também utilizasse ou definisse um padrão.

3 Interfaces entre Sistemas

3.1 XML

Atualmente nas cadeias produtivas, destacam-se as soluções derivadas de EDI (*Electronic Data Interchange*) (WANG; ZHANG, 2005) e XML, mas as tendências têm mostrado a utilização de sistemas via web com adição de semântica aos dados.

Para haver um livre intercâmbio de dados de um sistema para outro, é necessário que haja um padrão. Bio (2008) analisa esta situação:

Já quando se aborda a interface entre determinados sistemas, a questão é mais complexa. Inicialmente, devemos nos ater ao fato de que a grande maioria dessas interfaces dá-se especificamente com a base de dados do outro sistema e não com o “sistema” genericamente dito. Nestes casos, é preciso que haja

uma série de cuidados no sentido de atestar que as características dos dados do outro sistema sejam as mesmas requeridas pelo sistema em questão (e. g., temporalidade da renovação e atualização, dimensão, padrão de codificação, tipo físico de dado – alfanumérico, numérico, etc.). (BIO, 2008)

Na Figura 2a é apresentado um fluxo de informação sem a utilização de XML e a Figura 2b com XML. Na primeira figura é mostrada a interação de aplicações com informações extraídas de diversas fontes e formatos que são apresentadas também em diversas maneiras e formas indefinidas. Nesse processo, a utilização da informação é custosa e imprecisa devido ao excessivo manuseio para transformá-la em informação para a tomada de decisão.

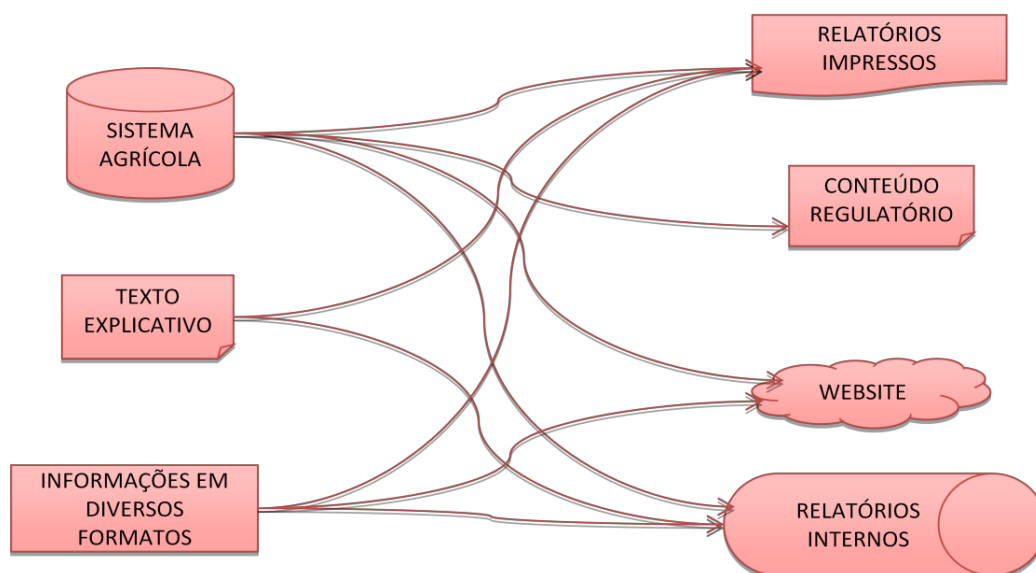


Figura 2a - Fluxo de informação SEM XML. Adaptado pelos autores de (RICCIO, 2005).

Já com o uso de XML (Figura 2b) o processo de transformação/divulgação dos dados acontece de forma padronizada. A conversão para o formato XML permite que a informação seja visualizada e personalizada conforme a necessidade. Nesse modelo também podemos destacar a possibilidade de análises automatizadas e atualizações mais frequentes da informação.

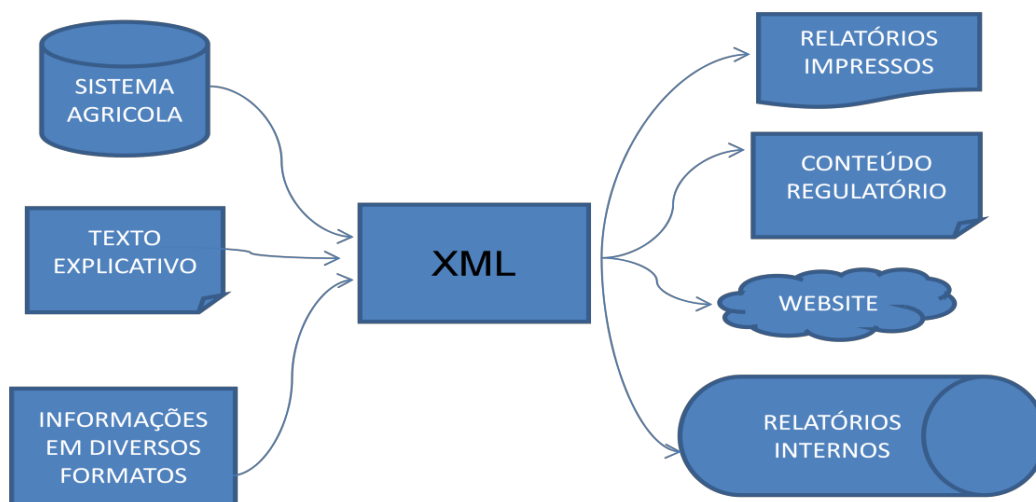


Figura 2b - Fluxo de informação COM XML. Adaptado pelos autores de (RICCIO, 2005).

Dentre os benefícios do uso de XML pode-se destacar a interoperabilidade entre os sistemas e plataformas de software. Aplicações podem ser desenvolvidas para ler de forma automática arquivos disponibilizados pelas organizações da cadeia produtiva e agregar valor nas tomadas de decisões. Por exemplo, a partir de uma base da UNICA (União da Indústria de Cana-de-Açúcar) que podem ser de distribuição automática – pode-se coletar informações de demonstrativos financeiros das usinas e gerar cubos de decisão e tabelas dinâmicas em sistemas de BI (*Business Intelligence*).

3.2 Padrões Agrícola: AgXML e AgroXML

Com base na XML diversos segmentos especificaram seus padrões que são aplicados a domínios específicos. Na agricultura existem diversos padrões, o que parece ser mais um problema. Muitos padrões para o mesmo domínio. No entanto há um esforço em todo mundo para a adoção de um mesmo padrão.

Na Europa uma das principais iniciativas é o AgroXML, um padrão para a troca de dados do setor agrícola desenvolvido e mantido pela KTBL (*The Association for Technology and Structures in Agriculture*) e parceiros na Alemanha. O AgroXML é baseado em XML e os termos agrícolas específicos são armazenados e explicados no

dicionário do AgroXML. O ponto forte dessa proposta é o fato de o esquema XML e o dicionário do AgroXML estar sendo submetido a um grupo para ser discutido e coordenado em uma força tarefa, na tentativa de obter o máximo de aceitação. Os potenciais usuários do AgroXML são os envolvidos na produção ou cadeia de suprimentos do setor agrícola europeu. Seu código é aberto e distribuído pelo consórcio W3C. A KTBL é responsável pela certificação e pelo *framework* chamado de agriXchange.

Nos Estados Unidos, empresas americanas em conjunto com o Departamento de agricultura participam do AgXML que define padrões para a troca de dados entre companhias de processamento de grãos e óleos de sementes que tem como objetivo criar padrões para o uso efetivo da comunicação de informações eletrônicas dentro da cadeia de suprimento do agronegócio como um todo (MURAKAMI, 2006). De forma geral, esses padrões buscam:

- Identificar os processos do negócio que, se fossem eletrônicos, melhorariam a eficiência e a eficácia do processo de negócio;
- Identificar os requisitos dos processos de negócio;
- Definir os esquemas XML e as diretrizes relacionadas que suportam os requisitos;
- Obter compromisso dos participantes na integração dos seus processos de negócio através da comunicação com mensagens baseadas em XML e na formação de um fórum para entendimento desses processos.

3.3 PAML

Precision Agriculture Markup Language (PAML) é uma iniciativa brasileira, desenvolvida no Laboratório de Automação Agrícola da USP por Murakami (2006). A PAML é uma linguagem padrão para troca de dados estruturados entre aplicações baseada na *eXtensible Markup Language* (XML). A definição do vocabulário dessa linguagem foi baseada no modelo de objetos para sistemas abertos de informação de campo (MOSAICo) (SARAIVA, 1998), no vocabulário controlado de terminologias da

SANTOS, F. de P.; MURAKAMI, E.

agricultura e domínios relacionados, AGROVOC (AGROVOC, 2006) e no modelo da *Geography Markup Language* (GML) (COX et al., 2002). GML é uma especificação em XML para transporte e armazenamento de informação geográfica, incluindo tanto propriedades espaciais quanto não espaciais de fenômenos geográficos.

A estrutura da PAML é definida por um esquema XML. *XML Schemas* (FALLSIDE; WALMSLEY, 2004), descrevem estruturas de documentos XML, definindo, por exemplo, os elementos e atributos que podem aparecer em um documento, os tipos de dados e valores padrão para os elementos e atributos, quais elementos são elementos filhos, etc. Com *XML Schema* é fácil descrever conteúdo de documentos, definir restrições e formatos de dados, validar dados e converter dados entre diferentes tipos. O esquema da PAML é uma extensão do esquema da GML.

A especificação GML define a sintaxe do seu esquema. A GML fornece uma estrutura aberta para a definição de esquemas de aplicações e objetos geográficos, suporta a descrição de esquemas de aplicações geográficas para domínios específicos, permite a criação e manutenção de esquemas e dados geográficos distribuídos e relacionados e fornece interoperabilidade entre Sistemas de Informação Geográfica (SIG).

Na Figura 3 é mostrado o diagrama de pacotes em UML dos esquemas que compõem os *frameworks* PAML e GML. Os pacotes destacados são os pacotes utilizados e testados por Murakami (2006).

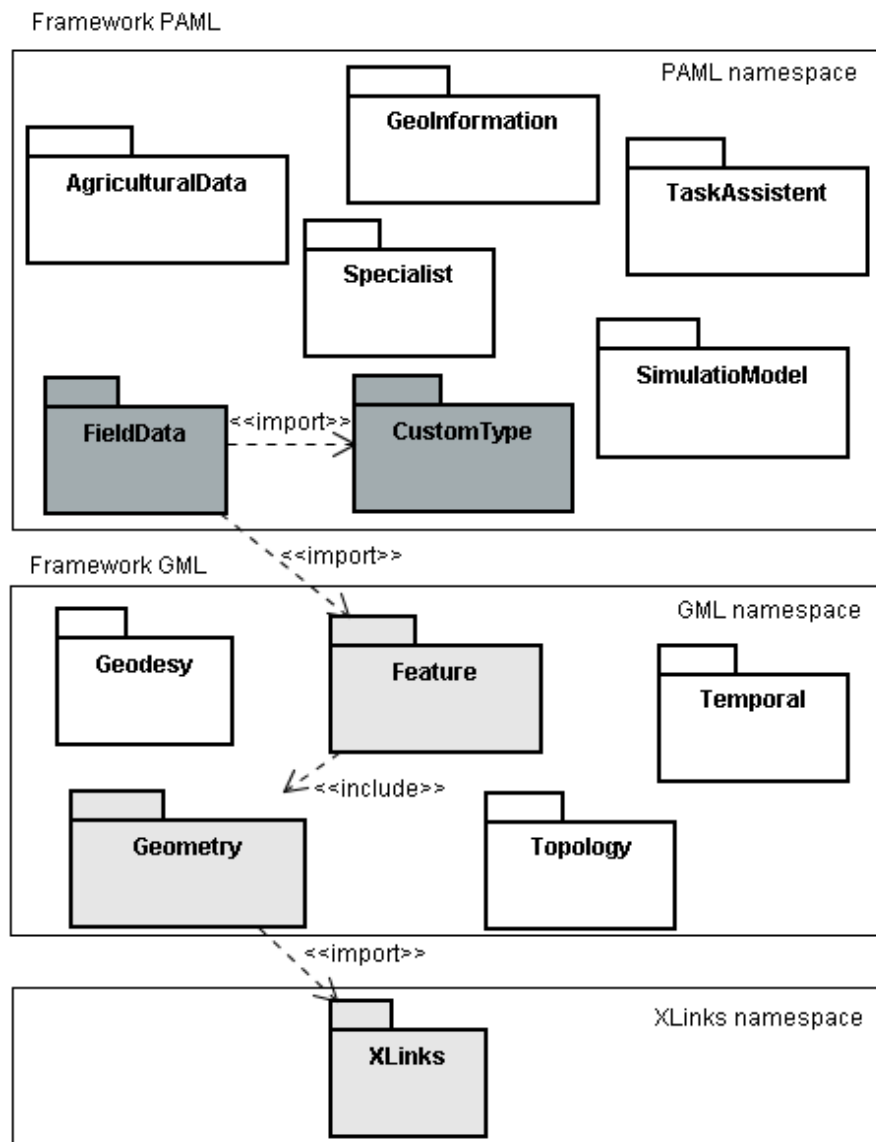


Figura 3 – Diagrama de dependência de pacotes em UML dos esquemas PAML e GML. (MURAKAMI et al., 2007)

Os esquemas base da GML, *Geometry*, *Feature* e *Xlink* podem ser vistos como os componentes de um *framework* de aplicação (*framework* GML) para desenvolvimento de esquemas que pertencem a um determinado domínio, por exemplo, agricultura de precisão (*framework* PAML). Os esquemas *FieldData*, *AgriculturalData*, *Specialist*, *TaskAssistent*, *GeoInformation* e *SimulationModel*, que correspondem aos subsistemas do MOSAICo, mais o *CustomType*, que são tipos de objetos e restrições reutilizáveis, formam o *framework*

PAML.

Na Figura 4 é mostrado o diagrama de classes do esquema *FieldData*, que estende as classes abstratas do esquema *Feature* da GML. Esta relação de herança garante compatibilidade com a GML.

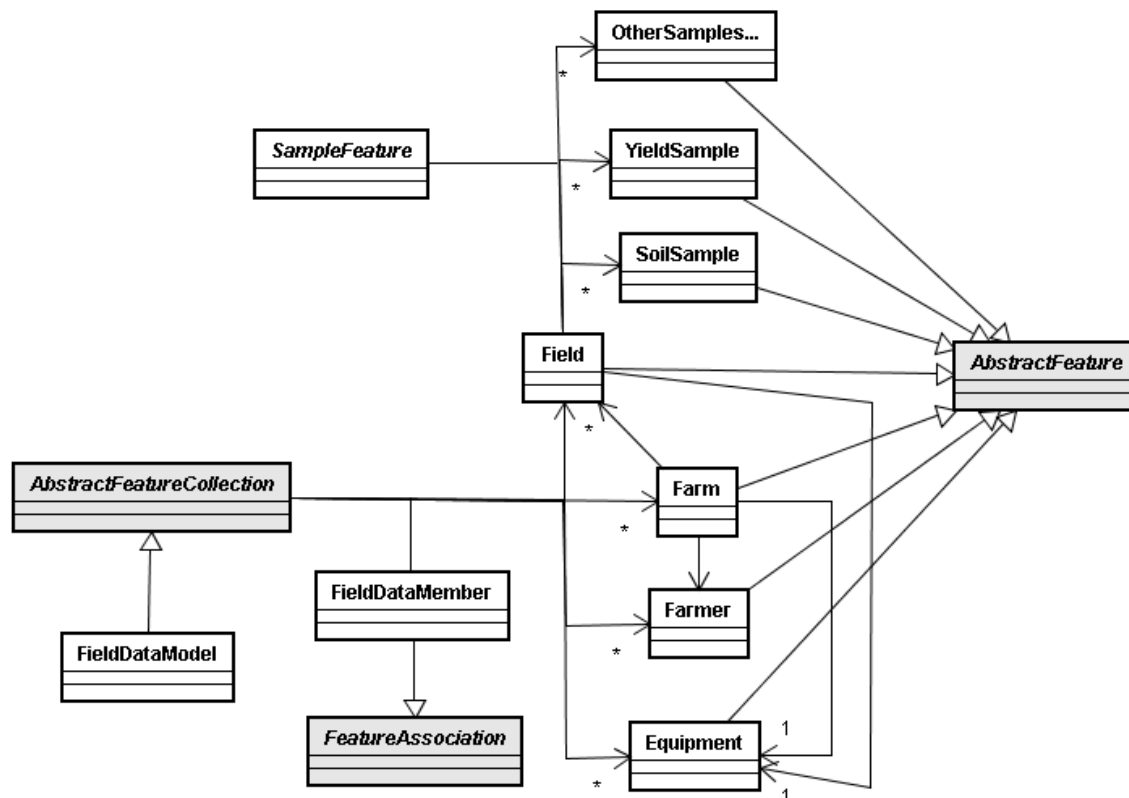


Figura 4 – Diagrama de classes do esquema *FieldData*, um *framework* de aplicação baseado no *framework* GML. (MURAKAMI et al., 2007)

Todas as classes que estendem de *AbstractFeature* são feições da PAML. Essas classes fazem parte do módulo Dados de Campo do MOSAICo. A classe *FieldDataModel* e *FieldDataMember* representam as coleções de feições da PAML. As classes destacadas em cinza são do esquema *Feature* da GML.

Na Figura 5 é mostrado um trecho do documento XML construído conforme definido no esquema da PAML, que corresponde a um documento a ser trocado entre aplicações. Este documento estrutura os dados de produtividade de grãos, o qual foi utilizado para integrar aplicações embarcadas em colhedadeiras, aplicações de filtragem de dados discrepantes e aplicações de geração de mapas aplicação de fertilizantes. Por

ser estruturado e validado por um esquema que define sua forma foi possível a manipulação automática dos dados pelos diversos sistemas envolvidos, na coleta, tratamento e aplicação.

```
<?xml version="1.0" encoding="UTF-8"?>
<FieldDataModel xmlns="http://www.padis.org"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.padis.org FieldData.xsd">

  <!-- AbstractFeatureType -->
  <gml:name>FieldDataModel-YieldSamples</gml:name>
  <fieldDataMember>
  ...
    <YieldSample>
      <id>1</id>
      <collectDate>2005-11-27</collectDate>
      <flow>60.0</flow>
      <cycles>60.0</cycles>
      <distance>60.0</distance>
      <swath>60.0</swath>
      <moisture>60.0</moisture>
      <field>60.0</field>
      <crop>60.0</crop>
      <altitude>60.0</altitude>
      <dryYield>60.0</dryYield>
      <easting>60.0</easting>
      <northing>60.0</northing>
      <pointDistance>60.0</pointDistance>
      <northDistance>60.0</northDistance>
      <eastDistance>60.0</eastDistance>
      <gml:Point
srsName="http://www.opengis.net/gml/srs/epsg.xml#4326">
        <gml:coord>
          <gml:X>20.0</gml:X>
          <gml:Y>5.0</gml:Y>
        </gml:coord>
      </gml:Point>
    </YieldSample>
  ...
    <dateCreated>2005-11</dateCreated>
</FieldDataModel>
```

Figura 5 – A feição YieldSample em um documento XML válido, conforme esquema FieldData.xsd (yieldsamples.xml). (MURAKAMI et al., 2007)

A PAML define a estrutura padrão para representação, transporte e armazenamento de dados. Por estender o *framework* GML, a PAML torna-se um *framework* de aplicação, que também pode ser estendido, sem perder a compatibilidade com a especificação GML, o que permite sua evolução sem impactar as aplicações que

SANTOS, F. de P.; MURAKAMI, E.

a utilizam como padrão para troca de dados.

Utilizando a mesma estratégia da criação da PAML, aplicada a Agricultura de Precisão, é possível definir uma linguagem padrão para troca de dados para o domínio da cana-de-açúcar.

4 WEB Services

Murakami (2006) propôs o uso de serviços web (*Web Services*) integrados a um barramento de serviços que utiliza documentos XML padronizados (PAML) para integração de aplicações. Ele destaca a independência de tecnologia e a padronização do formato de dados que a solução oferece. Apesar de ter sido aplicada ao domínio da agricultura de precisão, especificamente a grãos, a solução poderia ser aplicada facilmente no domínio da cana-de-açúcar.

Serviço web pode ser definido como um sistema de software projetado para suportar interação entre máquinas sobre uma rede. São inerentemente neutros de tecnologia, o que permite utilizá-los para integrar aplicações escritas em linguagens e paradigmas distintos, além de poderem ser perfeitamente utilizados para integrar sistemas legados, planilhas, banco de dados e outros serviços de diferentes provedores.

As informações geradas pela tecnologia da informação, sistemas legados, ERPs e outras fontes (planilhas eletrônicas, arquivos texto, até mesmo documentos XML) são disponibilizados na forma de serviços *web* (*Web Services*), o que torna possível que diversas aplicações clientes possam obtê-las e processá-las independentemente da tecnologia utilizada.

Porém, essa forma de integração de aplicações somente será econômica e tecnicamente possível através do uso de técnicas e conceitos computacionais modernos e de sistemas bem projetados, orientados por uma arquitetura capaz de fornecer propriedades como extensibilidade, reusabilidade e portabilidade.

A Arquitetura Orientada a Serviços (SOA) (PASLEY, 2005) tem sido utilizada para criação de componentes de negócio extensíveis e reusáveis com alta capacidade de integração. Nessa arquitetura, componentes são disponibilizados como serviços, com

contratos e interfaces bem definidas. Aplicações projetadas usando SOA podem ser incorporadas em uma aplicação composta mais facilmente que em uma aplicação monoliticamente projetada. Dois dos principais princípios dos componentes são facilmente obtidos com SOA, modularidade (divisão de grandes tarefas em tarefas menores) e encapsulamento (uso de interfaces claramente definidas para esconder as partes internas de um componente).

A abordagem orientada a serviços pode trazer diversos benefícios para os sistemas de informação para a indústria de cana-de-açúcar, como trás para a AP, dada a necessidade de integração com diversas fontes de dados e aplicações de negócios diferentes. Os benefícios dessa abordagem incluem:

- Desenvolvimento de componentes de negócio simplificados;
- Facilidade de montar soluções de negócios construídos como redes de serviços;
- Facilidade de integração de aplicações heterogêneas;
- Flexibilidade e agilidade nas mudanças;
- Proteção dos componentes de negócio devido a mudanças na tecnologia.

5 Conclusão

A indústria da informação está em constante evolução e, como ondas, assistimos a mudança da computação centralizada para as redes distribuídas; tecnologia cliente-servidor, computação na web e junto com elas, novas práticas de negócios e novas necessidades foram surgindo. À medida que as empresas do agronegócio distribuem suas informações pela internet, cresce a necessidade de troca de informações entre elas.

Neste artigo, de forma panorâmica, foram apresentados padrões de metadados para troca de informações automática entre aplicações. Uma solução baseada em serviços web que utiliza documentos XML padronizados como mensagens também foi apresentada, para mostrar que aplicações heterogêneas podem se comunicar, aproveitando infraestruturas computacionais existentes.

Portanto, o uso de serviços web com padrões de metadados baseados na

SANTOS, F. de P.; MURAKAMI, E.

linguagem XML pode reduzir o problema de integração dos sistemas legados, arquivos texto, planilhas eletrônicas, etc. e permitir a convivência harmônica desses sistemas com os novos ERP's especializados em uso atualmente na gestão da produção canavieira.

6 Referências Bibliográficas

ABRAHÃO, E., September 2009. An extension of the agroxml for the sugar cane crop in brazil. In: **Proceedings**. CIGR, International Commission of Agricultural and Biological Engineers, Argentina.

AGRIS AP. **Agris Application Profile: Methodology**. Disponível em: <www.fao.org/agris/agmes/Documents/ap_methodology.doc>. Acesso em: 27 fev. 2006.

AGROVOC. **Agrovoc Thesaurus**. Food and Agriculture Organization of the United Nations. Disponível em: <<http://aims.fao.org/website/AGROVOC-Thesaurus/sub>>. Acesso em: 26 jan 2012.

AgXML. **AgXML Standard**. AgXML L.L.C. 2004. Disponível em: <<http://www.agxml.org/Standard.html>>. Acesso em: 27 fev. 2006.

BERGERON, B. **Essencial of XBRL: financial reporting in the 21st century**. New Jersey: John Wiley & Sons, 2003. 220 p.

BIO, S. R. **Sistemas de Informação: um enfoque gerencial**. 2. ed. São Paulo: Atlas, 2008. 235 p.

COX, S. et al. **GML Implementation Specification**, version 2.1.2. OpenGIS Implementation Specification, 2002. Disponível em: <<http://www.opengeospatial.org/docs/02-069.pdf>>. Acesso em 27 fev 2006.

DOLUSCHITZ, R. et al. AgroXML – A standardized data format for information flow in agriculture. In: EFITA/WCCA JOINT CONGRESS ON IT IN AGRICULTURE, 2005, Portugal. **Proceedings**. Vila Real, 2005. p.439-443.

FALLSIDE, D.C.; WALMSLEY, P. **XML Schema Part 0: Primer Second Edition**, 2004. W3C Recommendation. Disponível em: <<http://www.w3.org/TR/xmlschema-0>>. Acesso em: 27 de fevereiro de 2006.

MURAKAMI, E. **Uma infra-estrutura de desenvolvimento de sistemas de informação orientados a serviços distribuídos para agricultura de precisão**. 2006. 192p. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2006.

MURAKAMI, E.; SARAIVA, A. M.; RIBEIRO JR., L. C. M.; CUGNASCA, C. E.; HIRAKAWA, A. R., CORREA, P. L. P. An infrastructure for the development of distributed service-oriented information systems for precision agriculture. **Computer Electronics Agriculture**, 2007

PASLEY, J. How BPEL and SOA are changing Web Services development. **IEEE Internet Computing**, v.9, n.3, p.60-67. 2005.

RICCIO, E. L. et al. (Org.). **A divulgação de Informações Empresariais XBRL: Extensible Business Reporting Language**. Rio de Janeiro: Editora Ciência Moderna, 2005.

SARAIVA, A.M. **Um Modelo de Objetos para Sistemas Abertos de Informações de**

Campo para Agricultura de Precisão – MOSAICo. 1998. 235p. Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 1998.

WANG, M. e ZHANG, S. Integrating EDI with an E-SCM system using EAI technology. **Information Systems Management**, vol. 22, Nº3, 2005. p. 31-37.