

**Análise de uso de padrões de metadados em projetos de pesquisa e desenvolvimento na Embrapa Informática Agropecuária**





*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Informática Agropecuária  
Ministério da Agricultura, Pecuária e Abastecimento*

# ***Boletim de Pesquisa e Desenvolvimento 27***

## **Análise de uso de padrões de metadados em projetos de pesquisa e desenvolvimento na Embrapa Informática Agropecuária**

*Carla Geovana do Nascimento Macário  
Eduardo Antonio Speranza  
Ivo Pierozzi Júnior  
Leonardo Ribeiro Queirós  
Marcos Cezar Visoli  
Roberto Hiroshi Higa  
Sílvia Maria Fonseca Silveira Massruhá*

Campinas, SP  
2011

## **Embrapa Informática Agropecuária**

Av. André Tosello, 209 - Barão Geraldo  
Caixa Postal 6041 - 13083-886 - Campinas, SP  
Fone: (19) 3211-5700 - Fax: (19) 3211-5754  
www.cnptia.embrapa.br  
sac@cnptia.embrapa.br

## **Comitê de Publicações**

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa, Stanley Robson de Medeiros Oliveira, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo, Paula Regina Kuser Falcão*

Supervisor editorial: *Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica: *Neide Makiko Furukawa*

Secretária: *Carla Cristiane Osawa*

## **1ª edição on-line 2011**

### **Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

### **Dados Internacionais de Catalogação na Publicação (CIP) Embrapa Informática Agropecuária**

---

Análise de uso de padrões de metadados em projetos de pesquisa e desenvolvimento na Embrapa Informática Agropecuária / Carla Geovana do Nascimento Macário... [et al.]. - Campinas : Embrapa Informática Agropecuária, 2011.  
23 p. : il. - (Boletim de pesquisa e desenvolvimento / Embrapa Informática Agropecuária, ISSN 1677-9266; 27)

1. Metadados. 2. Interoperabilidade. 3. Padrões de metadados.  
I. Macário, Carla Geovana do Nascimento. II. Embrapa Informática Agropecuária. III. Título. IV. Série.

---

CDD 025.04 (21. ed.)

© Embrapa 2011

# Sumário

<b>Resumo .....</b>	<b>5</b>
<b>Abstract.....</b>	<b>7</b>
<b>1 Introdução .....</b>	<b>8</b>
<b>2 Objetivo.....</b>	<b>9</b>
<b>3 Conceitos básicos .....</b>	<b>9</b>
<b>5 Aplicação e análise dos questionários.....</b>	<b>16</b>
<b>6 Conclusões.....</b>	<b>21</b>
<b>Referências .....</b>	<b>23</b>



# Análise de uso de padrões de metadados em projetos de pesquisa e desenvolvimento na Embrapa Informática Agropecuária

---

*Carla Geovana do Nascimento Macário<sup>1</sup>*

*Eduardo Antonio Speranza<sup>2</sup>*

*Ivo Pierozzi Júnior<sup>3</sup>*

*Leonardo Ribeiro Queirós<sup>4</sup>*

*Marcos Cezar Visoli<sup>5</sup>*

*Roberto Hiroshi Higa<sup>6</sup>*

*Sílvia Maria Fonseca Silveira Massruhá<sup>7</sup>*

## Resumo

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) produz uma grande quantidade de dados como resultado das pesquisas que realiza. Os dados gerados abrangem diferentes domínios: solos, clima, coleções, dados de animais, dados bibliográficos, entre outros. Muitas vezes os projetos trocam ou reúsam a informação produzidas. Apesar disso, muitos deles

---

<sup>1</sup> *Doutora em Ciência da Computação, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo, 13083-886, Campinas, SP, carla@cnptia.embrapa.br*

<sup>2</sup> *Mestre em Engenharia Elétrica, Analista da Embrapa Informática Agropecuária, speranza@cnptia.embrapa.br*

<sup>3</sup> *Doutor em Biologia, Pesquisador da Embrapa Informática Agropecuária, ivo@cnptia.embrapa.br*

<sup>4</sup> *Doutor em Engenharia Agrícola, Analista da Embrapa Informática Agropecuária, lqueiros@cnptia.embrapa.br*

<sup>5</sup> *Mestre em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, visoli@cnptia.embrapa.br*

<sup>6</sup> *Doutor em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, roberto@cnptia.embrapa.br*

<sup>7</sup> *Doutora em Computação Aplicada, Pesquisadora da Embrapa Informática Agropecuária, silvia@cnptia.embrapa.br*

ainda são armazenados de diferentes formas e usando diferentes formatos, como planilhas, sistemas de banco de dados, papel, entre outros.

A necessidade ou possibilidade de integração/compartilhamento de informação entre esses sistemas ou mesmo com outras instituições de pesquisa, desencadeou ações para a incorporação de novas estruturas e conceitos aos sistemas desenvolvidos, no sentido de facilitar a interoperabilidade de dados. Uma abordagem para esse problema é a adoção de padrões de metadados bem estabelecidos para descrição desses dados.

Nesse sentido, a Embrapa Informática Agropecuária estabeleceu um grupo de trabalho cuja missão era identificar a situação de seus projetos de pesquisa quanto à adoção desses padrões. O grupo elaborou um conjunto de perguntas que foram apresentadas à maioria desses projetos, permitindo não apenas identificar a utilização de metadados, mas também descrever as principais observações e recomendações sobre esse tema. Este documento apresenta o trabalho realizado por esse grupo, bem como os resultados obtidos.

Termos para indexação: Metadados, interoperabilidade, padrões de metadados.



# Analysis of the metadata standards uses in research and development projects at Embrapa Agricultural Informatics

---

## Abstract

*The Brazilian Research Corporation (Embrapa) produces a huge amount of data as a result of its research projects. The generated data comprise several domains, such as information on soil, climate, animals and vegetation and sometimes projects need to export or to import these information among them. Beside this, much of these data are still stored using different formats, like data base systems, worksheet and papers.*

*The need for information integration and sharing among systems and research companies, started a set of actions to incorporate new structures and concepts to the developed systems, aiming at data interoperability. An approach for this is the adoption of well known metadata standards, established to describe these data.*

*In this sense, Embrapa Informatics Agriculture established a workgroup with a mission to identify the situation of its research projects, considering the adoption of those standards. This group prepared a set of questions, which were presented for the researchers, allowing them not only to identify the use of metadata in their projects, but also to describe the main issues and recommendations about this theme. This document present the work performed by that workgroup and also the obtained results.*

*Index terms: Metadata, interoperability, metadata standard.*

## 1 Introdução

A Embrapa produz uma grande quantidade de dados como resultado das pesquisas que realiza. Os dados gerados abrangem diferentes domínios da ciência: solos, clima, coleções, animais, dados bibliográficos, entre outros. O crescimento da importância da tecnologia de informação (TI) dentro da empresa tem levado à construção de sistemas de informação que viabilizam a aquisição, o tratamento, a organização e o armazenamento desses dados. A Embrapa Informática Agropecuária, unidade de pesquisa da empresa focada em TI, tem participado do projeto e do desenvolvimento de diversos desses sistemas. Além disso, o seu foco de pesquisa - TI - faz com que conheça também sistemas desenvolvidos pelas demais unidades. Esse conhecimento global dos sistemas de software existentes na Embrapa, levou à identificação da necessidade ou possibilidade de integração desses sistemas, ou mesmo do compartilhamento dos dados com outras instituições de pesquisa e com a sociedade. Tal fato desencadeou ações para a incorporação de novas estruturas e conceitos aos sistemas desenvolvidos, no sentido de viabilizar a interoperabilidade de dados.

Os sistemas, em geral, têm seus dados armazenados de forma heterogênea, usando diferentes recursos, como planilhas eletrônicas, sistemas gerenciadores de banco de dados, catálogos de papel, entre outros. Essa heterogeneidade é comum no desenvolvimento de sistemas para atender a demandas específicas, geralmente num momento em que a preocupação com integração e compartilhamento não é tão evidente.

Para uniformizar o processo de intercâmbio de dados entre sistemas, diversas associações ou comitês têm sido criados com o propósito de definição de um conjunto de padrões, nacionais e internacionais, para descrição dos dados, facilitando o entendimento da semântica dos dados e a sua integração. Em geral, essa descrição é feita por meio de metadados e de mecanismos de comunicação entre os diferentes metadados.

Uma empresa como a Embrapa, que gera dados e conhecimento, deve se preocupar com isso, definindo e indicando o tipo/formato de informação a ser armazenada e disponibilizada. A Embrapa Informática Agropecuária vem envidando esforços nessa direção, tendo constituído um grupo de trabalho com o objetivo de levantar padrões de dados em uso na unidade.

Essa atividade é o marco inicial de um conjunto de ações que visam à elaboração de um plano guia para prover interoperabilidade de dados entre sistemas da unidade, que poderão futuramente ser exportadas para as demais unidades da empresa como forma de promover o compartilhamento dos diferentes tipos de dados existentes na empresa.

Este documento apresenta o trabalho realizado, que deve servir como base para atividades futuras relacionadas ao tema “interoperabilidade de dados”. A seção 2 descreve o objetivo do trabalho e a seção 3 apresenta conceitos básicos relacionados à interoperabilidade de dados, fornecendo ao leitor uma visão geral do tema. A metodologia de trabalho aplicada é descrita na seção 4 e os resultados obtidos são apresentados na seção 5, e comentados. Finalmente, a seção 6 conclui os trabalhos, com a sugestão de ações futuras a serem desenvolvidas. A lista de projetos entrevistados, bem como o questionário aplicado, compõem o anexo deste documento.

## **2 Objetivo**

Realizar levantamento dos padrões e tipos de dados utilizados nos projetos liderados pela Embrapa Informática Agropecuária, como passo inicial para o estabelecimento de um guia para prover interoperabilidade de dados nos sistemas de informação da Embrapa.

O grupo de trabalho focou na identificação do estado da arte da unidade quanto à promoção de interoperabilidade entre os sistemas de informação e bancos de dados desenvolvidos ou em uso nesses projetos.

## **3 Conceitos básicos**

Esta seção apresenta conceitos básicos relacionados ao tema interoperabilidade de dados, visando fornecer ao leitor uma visão geral do tema.

## 3.1 Interoperabilidade

(retirado de e-ping - Padrões de Interoperabilidade de Governo Eletrônico<sup>1</sup>)

### O que é interoperabilidade?

“Interoperabilidade não é somente integração de sistemas nem somente integração de redes. Não referencia unicamente troca de dados entre sistemas e não contempla simplesmente definição de tecnologia.

É, na verdade, a soma de todos esses fatores, considerando, também, a existência de um legado de sistemas, de plataformas de hardware e software instaladas. Parte de princípios que tratam da diversidade de componentes, com a utilização de produtos diversos de fornecedores distintos. Tem, por meta, a consideração de todos os fatores para que os sistemas possam atuar cooperativamente, fixando as normas, as políticas e os padrões necessários para consecução desses objetivos.

Para que se conquiste a interoperabilidade, as pessoas devem estar engajadas num esforço contínuo para assegurar que sistemas, processos e culturas de uma organização sejam gerenciados e direcionados para maximizar oportunidades de troca e reúso de informações, interna e externamente ao governo federal.”

## 3.2 Metadados

### 3.2.1 O que são metadados?

Metadados são, normalmente, definidos como dados que descrevem dados. Um metadado consiste em um campo e uma descrição associada. Por exemplo, numa biblioteca, título, autor e data de publicação são

---

<sup>8</sup> Disponível em: <<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padres-de-interoperabilidade>>

campos de metadados. A informação associada é usada para descrever o material cadastrado. Metadados são usados para:

- descrever um recurso para sua descoberta e identificação;
- descrever a organização de um recurso;
- fornecer informação de direitos autorais e restrições de uso;
- fornecer informações administrativas desses dados, como data de criação, autoria;
- preservar informação de dados;
- agregar outras informações a um dado, como aquelas provenientes de uma análise;
- organizar informações a serem disponibilizadas em um portal;
- permitir que um dado seja entendido por humanos e máquinas, facilitando o seu intercâmbio entre sistemas com diferentes hardware e software, estruturas de dados e interfaces.

Numa empresa, os metadados são usados para descrever os recursos dos seus dados e a forma como estão organizados. Com isso permitem a divulgação da informação disponível, permitindo que usuários desses recursos conheçam os dados existentes. Além disso, o uso de metadados aumenta as oportunidades de interoperabilidade, tornando múltiplas fontes de informação acessíveis a vários consumidores, preservando os dados. Metadados e servidores de metadados têm papel crucial na integração de dados, permitindo que usuários integrem dados de múltiplas fontes, organizações e formatos.

A importância de metadados e a necessidade cada vez mais presente de interoperabilidade levou ao desenvolvimento de iniciativas para obtenção de metadados específicos para diferentes domínios, considerando os diferentes formatos, comunidades e agência de dados. A adoção desses padrões é o passo inicial para interoperabilidade.

### **3.2.2 Existem padrões de metadados?**

A adoção de padrões de metadados para intercâmbio entre projetos e instituições brasileiras vem crescendo. Padrões de metadados também vêm sendo adotados por governos em diversos países.

A seguir são apresentados exemplos de padrões de metadados em uso, organizados por alguns domínios de informação que a Embrapa Informática Agropecuária trabalha. Uma lista categorizada mais completa pode ser uma atividade futura a ser realizada na unidade, conforme sugestões descritas na seção final desse relatório.

## **Dados geográficos**

Dados geográficos, ou geoespaciais, são aqueles que fazem referência a uma localidade na superfície da Terra. Exemplos desses dados são imagens de satélite, mapas georreferências e séries temporais com dados de temperatura.

Dados geográficos podem ser descritos usando dois padrões já bem estabelecidos: o ISO 19115 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008), um padrão europeu, e o FGDC (UNITED STATES, 1998), um padrão americano. O objetivo desses padrões é prover um conjunto de definições e terminologias comuns para a documentação e a troca de dados geográficos.

O padrão ISO 19115 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008) é um padrão de informação geográfica que define o esquema requerido para descrever informação e serviços. Provê informação sobre a identificação do dado, extensão, qualidade, esquema espacial e temporal, referências espaciais e distribuição geográfica. No total são 326 campos de metadados organizados em 2 classes (SILVA, 2008).

O padrão Federal Geographic Data Committee<sup>1</sup> (FGDC) implementa a Infraestrutura Americana de Dados Espaciais (NSDI). O Padrão para Metadados Geoespaciais Digitais (CSDGM), geralmente referenciado como o padrão de Metadados FGDC, permite a definição de perfis e é extensível por definições de seus usuários. Todas as agências americanas adotam esse padrão para documentar seus dados espaciais. O padrão consiste de sete seções principais de metadados e três seções de apoio. As seções principais contêm informação de identificação, qualidade de dados, organização e referência espacial e distribuição. As seções de apoio contêm informação de citação, tempo e contato (ESTADOS UNIDOS, 1998).

O Perfil de Metadados Geoespaciais do Brasil (MGB) (BRASIL, 2011) é um perfil de metadados baseado na norma ISO 19115 e criado com o objetivo de se obter um subconjunto mínimo de metadados para descrever dados geoespaciais gerados no país. Foi elaborado pela Confederação Nacional de Cartográfica (Concar), órgão colegiado do Ministério do Planejamento, Orçamento e Gestão (MPOG). Dentre outras coisas, a Concar propôs, em 2007, o estabelecimento da Infraestrutura Nacional de Dados Espaciais (Inde), composta por um conjunto integrado de políticas, arranjos institucionais, padrões, recursos humanos e tecnológicos, necessários para facilitar a disponibilização, o acesso e o uso de dados e informações geoespaciais. Com isso, pretende-se catalogar, integrar e harmonizar dados geoespaciais existentes nas instituições do governo brasileiro, produtoras e mantenedoras desse tipo de dado, de maneira a facilitar sua localização, exploração e acesso.

### **Dados de biodiversidade**

Um padrão bastante usado na troca de dados de biodiversidade é o Darwin Core2. Esse padrão é baseado no padrão Dublin Core3 e visa facilitar o intercâmbio de informação sobre a ocorrência geográfica de espécies e a existência de espécimes em coleções. Inclui metadados para a especificação do nome do organismo, local, data e autor de uma coleta de espécie.

Já o projeto Global Biodiversity Information Facility (GBIF)4 tem como missão disponibilizar os dados mundiais da biodiversidade, de forma universal, livre e via internet. Para isso, o projeto pretende prover uma infraestrutura de informática essencial para a pesquisa em biodiversidade, como padrões de metadados para a representação e intercâmbio, protocolos, serviços e ferramentas que os implementem.

### **Dados bibliográficos**

O padrão Machine-Readable Cataloging (MARC) (LIBRARY OF CONGRESS, 2011a, 2011), foi proposto em 1960 pela Biblioteca do Congresso como um formato para descrição de material bibliográfico. Ele é amplamente adotado, servindo como base para outros padrões, usado para descrever catálogos de “on-line” de acesso público. Um deles é o

MARC21, resultado da combinação do formato MARC americano e canadense.

Um registro MARC é composto por três elementos: estrutura, indicação do conteúdo e conteúdo propriamente dito. A estrutura do registro é uma implementação dos padrões internacionais ANSI Z39.2 e ISO 2709. As indicações de conteúdo são códigos e convenções estabelecidos para identificar e caracterizar os dados dentro do registro e permitir sua manipulação. Os conteúdos dos dados que compõem um registro MARC, geralmente, são definidos por padrões externos ao formato, como: International Standard Bibliographic Description (ISBD), Anglo-American Cataloguing Rules (AACR2), Library of Congress Subject Headings (LCSH) ou outros códigos usados pela instituição criadora do registro.

O Metadata Object Description Schema (MODS) (LIBRARY OF CONGRESS, 2011b) é um esquema de metadados derivado do MARC21, desenvolvido para descrever dados diversos e, particularmente, formatos bibliográficos. Inclui um subconjunto de campos do padrão MARC e, em alguns casos, agrupa elementos do formato bibliográfico MARC 21. Dentre suas vantagens destacam-se: a) o propósito de complementar outros formatos de metadados; b) ser mais rico que o formato Dublin Core; c) ser mais simples de usar e mais amigável que o formato MARC.

## **Recursos digitais**

O padrão Dublin Core (DUBLIN CORE METADATA INITIATIVE, 2011) é um esquema de metadados que foi desenvolvido visando à descrição de recursos digitais de diferentes áreas, como vídeos, software, imagens, sites na web e serviços. Esse padrão foi proposto pela Dublin Core Metadata Initiative (DCMI), uma organização dedicada a promover a adoção de padrões de interoperabilidade e desenvolver vocabulários especializados para descrever fontes que tornem mais inteligentes os sistemas de descobrimento de informações.

Características como simplicidade e concisão fazem com que o Dublin Core seja largamente utilizado.



## Metodologia de trabalho

A metodologia de trabalho adotada pelo grupo divide-se em 3 conjuntos de ações, a saber: ações motivadoras, orientadoras e executoras.

As ações motivadoras tinham como objetivos motivar a equipe técnica da Embrapa Informática Agropecuária a se familiarizar com os conceitos de interoperabilidade e padrões de dados. Nesse sentido, dois eventos foram realizados:

- condução de uma discussão sobre os conceitos relacionados à interoperabilidade no Blog do café científico, a partir do post “Dados, Informação, Conhecimento: Inovação – Novas tecnologias – Cloud Computing – Agricultura sustentável” do dia 16 de julho - (<http://www.cc.cnptia.embrapa.br/?p=86>).
- realização de um Café Científico (presencial) no dia 27 de agosto, no qual o tema interoperabilidade foi discutido com grande parte da equipe técnica do Centro.

Em paralelo às ações motivacionais, foram realizadas reuniões com os membros da comissão, visando direcionar o trabalho do grupo. Nessas reuniões, decidiu-se pela realização de um levantamento formal de questões relacionadas à interoperabilidade entre os projetos em andamento e sob liderança da Embrapa Informática Agropecuária. Para isso, as seguintes atividades foram desenvolvidas:

- levantamento, junto à Secretaria Executiva do CTI, da lista de projetos em andamento na unidade e seus respectivos líderes (vide Anexo I).
- elaboração de questionário para encaminhamento para os líderes de projetos (vide Anexo II).

Novamente, visando facilitar a familiarização da equipe técnica da unidade com os conceitos de interoperabilidade, também foi enviado, para os líderes de projeto, um documento trazendo um resumo dos principais conceitos sobre o tema (conforme apresentado na seção 2 deste documento).

Após receber os questionários respondidos, eles foram tabulados e analisados, conforme apresentado na seção 5.

Ressalta-se, aqui, que a carteira de projetos pode não retratar a realidade da unidade por completo, visto existirem sistemas de informação ou bancos de dados em desenvolvimento, em planos de ação de projetos liderados por outras unidades, ou que foram desenvolvidos em projetos já finalizados.

## 5 Aplicação e análise dos questionários

Esta seção descreve a aplicação dos questionários e a tabulação dos resultados, comentando-os.

### 5.1 Metodologia para aplicação dos questionários

A aplicação dos questionários e a tabulação dos resultados obtidos foi realizada em três fases. descritas a seguir: aplicação, tabulação e classificação.

#### **Fase 1: aplicação dos questionários**

Os questionários foram distribuídos aos líderes de projetos, que tiveram um prazo de duas semanas para resposta e devolução (houve tolerância de 2 dias de atraso na devolução). Apenas os questionários devolvidos no prazo foram considerados na análise.

#### **Fase 2: tabulação dos resultados**

Para tabulação dos questionários, consideraram-se apenas as questões 2, 4, 6, 7, 8 e 11, por serem aquelas que abordam diretamente a questão de interoperabilidade e de formatos de dados. Foi obtida a média de valores para cada questão, permitindo uma visão geral da situação dos projetos analisados. Para essa análise, foi considerada a seguinte escala:

- cada resposta N foi tabulada como -1 e cada resposta S como +1;

- valores faltantes (*missing*) foram desconsiderados para obtenção da média;
- escala considerada para categorização:
  - -1 a -0.7499: Forte N;
  - -0.75 a -0.2499: Moderado N;
  - -0.25 a 0.2499: Indefinido;
  - 0.25 a 0.7499: Moderado S;
  - 0.75 a 1: Forte S.

### **Fase 3: classificação dos projetos quanto à interoperabilidade**

Uma vez tabulados os questionários, procedeu-se à classificação dos projetos quanto a ações de interoperabilidade. Para isso, valores faltantes (*missing*) foram considerados como 0 (zero).

De posse dessa classificação, foi obtida nova média das respostas das 2, 4, 6, 7, 8 e 11, visando obter um retrato da situação de cada grupo em relação à interoperabilidade.

## **5.2 Resultados consolidados**

- 64% (21 em 33) foram respondidos.
- Apenas um projeto respondeu não para a questão 1, ou seja, não está sendo considerado repositório de dados no projeto. Esse projeto foi desconsiderado no restante da análise.
- Cinco projetos responderam N para a questão 2, sendo que, destes, um não respondeu a questão 3.
- Quatro projetos responderam N para a questão 4, sendo que, desses dois, responderam N para a questão 5 e dois responderam S.
- Dez projetos responderam N para a questão 8, sendo que sete responderam S para a questão 9 e um respondeu N.
- Dos cinco projetos que responderam S para a questão 8, quatro responderam S para a questão 10 e dois responderam N.
- Treze projetos indicam utilizar algum formato de dados e/ou padrão de metadados.
- Os formatos de dados e padrões de metadados adotados são: MGB/INDE, COSAVE, AIE, mmCIF, textual PDB, Dublin Core, shapefiles e

GeoTiff (geoespaciais), FASTA e GFF (bioinformática) e banco de dados relacional<sup>9</sup>.

- Treze projetos adotam um formato para importação/exportação dos dados, sendo vários específicos da área de conhecimento do projeto, como é o exemplo dos projetos de bioinformática ou biologia computacional; outros formatos adotados citados são o CSV e o XML.
- Cinco projetos indicam a utilização de vocabulários controlados, os quais convergem basicamente para os thesaurus da área de agricultura (Thesagro, AGROVOC e NAL) e para a Gene Ontology, no caso dos projetos de bioinformática.
- Em relação a padrões de metadados, apenas seis projetos indicam sua utilização para facilitar a interoperabilidade entre aplicações e citam o Dublin Core e outros específicos de cada área de conhecimento, principalmente bioinformática/biologia computacional;
- Oito projetos planejam adotar algum padrão de metadados, mas apenas dois projetos conseguem explicitar os padrões pretendidos: MGB (INDE) e ISO19115, ambos para dados espaciais.
- A pergunta 11 é se o líder tem sugestões a acrescentar. Ela foi respondida como um forte S pelos líderes dos projetos do grupo 1.

### Tabulação dos resultados

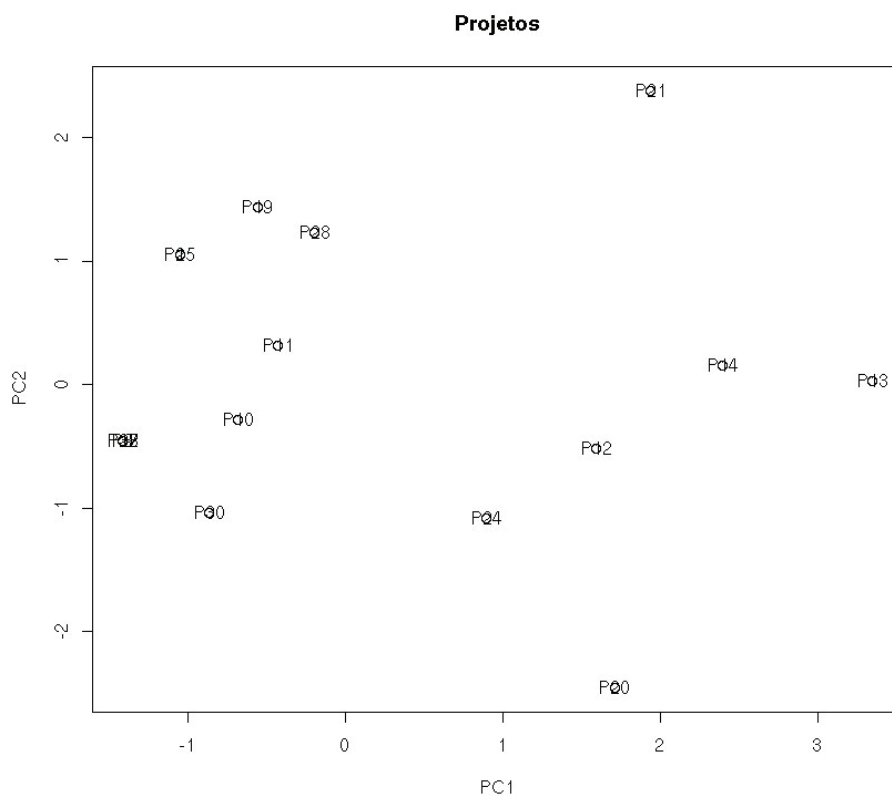
Considerando apenas as questões 2, 4, 6, 7, 8 e 11, obtém-se as seguintes médias e classificações indicadas na Tabela 1.

**Tabela 1.** Resposta geral para as questões Q2, Q4, Q6, Q7, Q8 e Q11.

	Q2	Q4	Q6	Q7	Q8	Q11
<b>Média</b>	0.4444	0.5556	0.5294	0.7333	-0.25	0.5
<b>Classe</b>	MS	MS	MS	MS	Indef	MS

Considerando os valores faltantes, é possível fazer uma análise de componentes principais, conforme ilustrado na Figura 1. A Figura 1 sugere que os projetos podem ser separados em dois grupos.

<sup>9</sup> Esta citação veio das respostas dos questionários. O grupo manteve todas, mesmo quando não se tratava de padrão ou formato, como no caso de "banco de dados relacional".



**Figura 1.** Os 2 componentes principais para os projetos, utilizando as respostas como variáveis.

Utilizando k-means e os componentes principais para dividir os projetos em dois grupos obtém-se:

**Tabela 2.** Grupos obtidos por k-means.

Grupo 1	Grupo 2
P8, P9, P10, P11, P15, P17, P19, P22, P25, P28, P30 e P33	P12, P13, P14, P20, P21 e P24

As Tabelas 3 e 4 apresentam as respostas para cada grupo, utilizando o mesmo processo anteriormente utilizado para avaliação das respostas de todos os projetos (Tabela 1).

**Tabela 3.** Resposta do grupo 1 para as questões Q2, Q4, Q6, Q7, Q8 e Q11.

	Q2	Q4	Q6	Q7	Q8	Q11
Média	0.8333	1	1	1	-0.2727	0.8182
Classe	S	S	S	S	Indef	S

**Tabela 4.** Resposta do grupo 2 para as questões Q2, Q4, Q6, Q7, Q8 e Q11.

	Q2	Q4	Q6	Q7	Q8	Q11
Média	-0.3333	-0.3333	-0.6	0.2	-0.2	-0.2
Classe	MN	MN	MN	Indef	Indef	Indef

## 5.2 Comentários

Considerando os resultados obtidos, é possível afirmar que:

- dos projetos que não utilizam metadados, a maioria (7) planeja adotar um padrão;
- dos projetos que utilizam metadados, 4 também utilizam vocabulários controlados;
- existem dois grupos de projetos: a) um (grupo 1) que está já está trilhando o caminho de adoção de padrões e b) outro (grupo 2) que parece não atento a essas questões.
- é possível concluir que os projetos do Grupo 1 exportam e importam dados, utilizam formatos para armazenar esses dados, bem como no processo de importação/ exportação.
- mesmo no Grupo 1, a questão de metadados (Q8) ainda não está solidificada, provavelmente com alguns projetos fazendo seu uso, enquanto que outros sequer têm conhecimento do recurso;
- uma hipótese a ser verificada é se os projetos do Grupo 1 são os que estão mais próximos “interoperar” dados.

- existe uma necessidade de uniformização de conceitos; por exemplo, banco de dados relacional não é padrão ou formato de dados.

## 6 Conclusões

O grupo estabelecido teve, por objetivo, realizar levantamento dos padrões e tipos de dados utilizados nos projetos liderados pela Embrapa Informática Agropecuária, como passo inicial para o estabelecimento de um guia para prover interoperabilidade de dados nos sistemas de informação da Unidade. Nesse sentido, o grupo aplicou um questionário que permitisse identificar o estado da arte da unidade quanto à promoção de interoperabilidade entre os sistemas de informação e bancos de dados desenvolvidos ou em uso nesses projetos. Este documento apresenta os resultados obtidos com a aplicação desse questionários, comentando-os.

Vale destacar alguns pontos. Optou-se pela aplicação do questionários considerando apenas os projetos sob liderança dessa unidade. Visto que existem sistemas de informação ou bancos de dados que vêm sendo desenvolvidos em planos de ação de projetos liderados por outras unidades, ou que já tiveram seu projeto finalizado, é possível que os resultados obtidos não retratem a completa realidade da unidade. Mesmo assim, servem como indicativo da situação da Unidade em se tratando do tema interoperabilidade.

Os resultados obtidos mostram que existem dois grupos de projetos: a) um (grupo 1) que já está trilhando o caminho de adoção de padrões e b) outro (grupo 2) que parece não atento a essas questões. Além disso, o número expressivo de perguntas não respondidas indicam uma necessidade de mais esclarecimento e sensibilização sobre o tema dos padrões de dados, interoperabilidade e sobre os recursos disponíveis para serem aplicados aos contextos dos projetos da Unidade. Há uma indicação de que o interesse existe. Mas também há indicação de que muitos pesquisadores ainda não conhecem bem as implicações e benefícios da utilização dos padrões, dos metadados e dos vocabulários controlados.

Assim, o grupo conclui que a Unidade demonstra interesse e preocupação com o tema abordado. Portanto, é essencial o desenvolvimento de ações gerenciais que viabilizem e motivem a ampla adoção de padrões e serviços focando em interoperabilidade de dados. Neste sentido, o grupo apresenta algumas recomendações, descritas a seguir.

1. Levantar relação de Planos de Ação não liderados pela Unidade, bem como outras ações que geraram banco de dados e não responderam ao questionário; enviar questionário aos responsáveis.
2. Realizar novo esforço junto aos líderes dos projetos que não responderam ao questionário, para que o façam.
3. Caso a Chefia decida direcionar esforços no sentido de ter um projeto que interopere dados, os candidatos mais viáveis estão no grupo 1. Recomenda-se, então, uma negociação com os respectivos líderes.
4. Analisar mais profundamente o que diferencia os grupos 1 e 2. São características dos projetos, das áreas, dos líderes?
5. A existência do grupo 2 pode indicar que alguns projetos não estão atentos para a questão de metadados nem para a questão de formatos de dados padrão, ou a sua adoção não é necessária. Nesse caso, é preciso estudar qual apoio/infraestrutura que precisam ser criados para alavancar tais projetos com relação a esse tema.
6. Ouvir os grupos identificados na recomendação 5 para levantar sugestões, visando melhorar o processo de adoção de padrões para interoperabilidade. Esse parece ser o caminho mais lógico para iniciar o estabelecimento de uma política de dados na Unidade.
7. Levantar sugestões junto ao grupo 1 para melhorar/facilitar/acelerar/aperfeiçoar o processo que eles já vêm trilhando.
8. Verificar os fatores que impedem o grupo 2 de trilhar o mesmo caminho e, supondo que a interoperabilidade seja uma questão de relevância para a Unidade, promover ações no sentido de compartilhar com os projetos do grupo 2 o conhecimento que os projetos do grupo 1 acumularam.



## Referências

BRASIL. Infraestrutura Nacional de Dados Espaciais. **Inde**. 2011. Disponível em: <<http://www.inde.gov.br/>>. Acesso em: 15 dez. 2011.

DUBLIN CORE METADATA INITIATIVE. **Dublin Core Metadata Initiative**. 2011. Disponível em: <<http://dublincore.org/>>. Acesso em: 15 dez. 2011.

ESTADOS UNIDOS. Federal Geographic Data Committee. **Content standard for digital geospatial metadata**. Washington, D.C., 1998. (FGDC-STD-001-1998).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19115**: 2003 Geographic information metadata. 2008. Disponível em: <<http://www.iso.org/iso/home.htm>>. Acesso em: 15 dez. 2011.

LIBRARY OF CONGRESS. **MARC - Machine Readable Cataloging**. 2011a. Disponível em: <<http://www.loc.gov/marc/bibliographic/>>. Acesso em: 15 dez. 2011.

\_\_\_\_\_. **MODS - Metadata Object Description Schema**. 2011b. Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 15 dez. 2011.

SILVA, H. MIG - **Metadados para Informação Geográfica - Introdução à Norma ISO 19115**. Portugal: Instituto Geográfico Português, Portugal, 2008. Slides. Disponível em: <[scrif.igeo.pt/webmig/docs/IntroducaoNorma19115.pdf](http://scrif.igeo.pt/webmig/docs/IntroducaoNorma19115.pdf)>.



---

*Informática Agropecuária*