

**MAPOTECA DIGITAL - UMA FERRAMENTA DE GESTÃO DO CONHECIMENTO  
BASEADA EM METADADOS**

Geraldo de Oliveira Santos, DSc. – Petrobras, geraldo.santos@petrobras.com.br

Jorge de Rezende, DSc. – [jrezende@gmail.com](mailto:jrezende@gmail.com)

Luiz Eduardo Seabra Varella – Petrobras, luizedu@petrobras.com.br

**RESUMO**

O objetivo deste trabalho é de analisar a tecnologia dos metadados como ferramenta de Gestão do Conhecimento. É apresentado, como exemplo, um modelo de dados georeferenciado de metadados, e um aplicativo, denominado Mapoteca Digital, que visa o gerenciamento e controle de documentos gráficos utilizados nas áreas de exploração e produção de petróleo da Petrobras.

Palavras-chave: Datamining, metadados

**1. Introdução**

O conceito de metadado não é novo, porém, o papel dos metadados e sua importância no ambiente de negócios certamente o é (MARCO, 2000). O metadado tem como objetivo permitir uma melhor integração, troca, acesso e interpretação das informações (GUNTHER, 1998). Os metadados estão presentes em várias áreas do conhecimento, tais como: informações geográficas, música, biblioteca digital, gestão do conhecimento, tecnologia da informação (Data Warehouse, e-bussiness, sistemas de apoio a decisão, Internet, agentes inteligentes, Crm, etc.), documentação eletrônica, processamento de imagens, educação, telecomunicações, entre outras.

Segundo SIMON, TOMASIC & GALHARD (1998), não existe uma definição concisa do que vem a ser metadado, apenas uma noção mais ou menos intuitiva que é “dado sobre o dado”.

O metadado é o dado que descreve a essência, os atributos e o contexto de um recurso (documento, fonte, sistema, imagem, regra, etc.) e caracteriza suas relações, visando seu acesso e utilização. O metadado pode ser considerado sinônimo de ponto de acesso, termo da área da catalogação. A finalidade principal dos metadados deve ser a de documentar e organizar, de forma estruturada, os dados de uma organização, com o objetivo de minimizar a duplicação de esforços e facilitar a manutenção dos mesmos TANNENBAUM (2002).

As pessoas que tem contato com alguma ferramenta de tecnologia da informação, utilizam alguma forma de metadado, mesmo sem saber que o usa ou qual é o significado. Isto é normal

até porque a sua própria definição não é um consenso, gerando alguma confusão. Encontram-se, na literatura, diferentes interpretações para metadado (SIMON, TOMASIC & GALHARD, 1998; TANNENBAUM, 2002; MARCO, 2000), tais como:

- Metadados são dados que descrevem atributos de um recurso, suportando algumas funções: localização, descoberta, documentação, avaliação, seleção, etc.
- Metadados fornecem o contexto para entender os dados com a passagem do tempo – conceito de versão;
- Metadado é a estrutura associada com os objetos, auxiliando seus usuários potenciais a ter o conhecimento da sua existência e/ou suas características;
- Metadados é um recurso que transforma o dado bruto em informação/conhecimento.

Para MORIATY (1997), as diferentes visões sobre metadados estão relacionadas, diretamente, ao estágio da organização ou instituição no contexto da gestão do conhecimento. Organizações com menor foco na gestão do conhecimento gerenciam dados brutos, e organizações mais focadas e direcionadas para a gestão do conhecimento são capazes de administrar seus recursos de informação e conhecimento de forma profunda.

A importância dos metadados pode ser vista da seguinte forma: o dado sem o metadado é como se fosse um turista em uma cidade desconhecida sem qualquer informação sobre a mesma. O turista (ou usuário) fica sem uma forma de orientação para obter a informação desejada.

Segundo MARCO (2000) e TANNENBAUM (2002) existem duas grandes categorias de metadados: os metadados técnicos e os metadados de negócio. O metadado técnico se refere à descrição dos dados necessários pelas várias ferramentas de tecnologia da informação para armazenar, manipular ou movimentar as informações (banco de dados, ferramentas CASE, ferramentas de modelagem, etc.). O metadado de negócio é a descrição das informações necessárias aos usuários de negócio, com a finalidade de prover o entendimento e o contexto da realidade de negócio e o significado das informações.

Os metadados vão adquirir cada vez mais importância na nova economia, principalmente com a integração entre a tecnologia da informação (Web, Data Warehouse, CRM, ERP, etc) e a gestão do conhecimento. Esta união resultará em um único ponto de acesso para a informação do

negócio através de um navegador de metadados. Os metadados se tornarão um componente crítico para qualquer arquitetura de negócio, TANNENBAUM (2002).

Por definição, a Gestão do Conhecimento é vista como sendo um conjunto de processos que agem sobre a criação, disseminação e utilização do conhecimento em toda a organização, com a finalidade de auferir vantagem competitiva, preferencialmente sustentável. Esses processos são classificados segundo a taxonomia CRIE/UFRJ em:

- **Gerar/Adquirir:** são os atos de prospectar, visualizar, avaliar, qualificar, triar, selecionar, filtrar, coletar, identificar. É onde o conhecimento surge interna ou externamente. Esta etapa procura capturar para a organização os conhecimentos necessários para obtenção de uma vantagem competitiva sustentável;
- **Organizar/Armacenar:** são os atos de explicitar, analisar, customizar, contextualizar e documentar. Referem-se ao armazenamento do conhecimento e a sua internalização. Tem por finalidade colocar o conhecimento em um formato utilizável, para que a empresa possa acessá-lo de maneira adequada.
- **Distribuir/Compartilhar:** são os atos de disseminar, dividir, distribuir. É a socialização do conhecimento. Refere-se a tarefa de notificar e tornar público os conhecimentos capturados e organizados, de forma clara e facilmente acessível para todos os funcionários e/ou colaboradores de uma organização.
- **Aplicar:** é o ato de produzir e usar. É onde o conhecimento agrega valor. Têm-se o conhecimento utilizado no dia-a-dia, principalmente apoiando decisões.
- **Criar:** é o ato de evoluir, inovar.

O metadado, por suas características, está fortemente envolvido com o Capital Estrutural. Com base na taxonomia acima conclui-se que o metadado está mais intensamente relacionado com o processo de organização do conhecimento, e medianamente relacionado com o processo de compartilhamento.

## 2. Padrões de metadados

São grandes os esforços para a definição de um formato comum para os metadados. Existem, publicados, diversos padrões de metadados, voltados para domínios particulares da ciência e da tecnologia. Entretanto, não existe nos ambientes científicos e de negócios um padrão comum capaz de atender toda e qualquer situação, devido à natureza heterogênea das aplicações científicas e tecnológicas. Os principais padrões de mercado são:

- CSDGM (Content Standards for Digital Geospatial Metadata): padrão de metadados estabelecido pelo Comitê de Dados Geográficos Federais dos Estados Unidos (FGDC – US Federal Geographic Data Committee), o qual serviu de modelo para o esforço de documentação de metadados do governo dos Estados Unidos (CLINTON, 1994). O padrão do FGDC é bastante complexo, consistindo de 334 elementos diferentes, sendo que desses, 119 são elementos compostos, ou seja, existem apenas para conter outros elementos. Entretanto, esses elementos compostos são importantes porque descrevem o relacionamento entre os outros elementos. Na realidade, o padrão do FGDC especifica o conteúdo, não especificando o formato dos arquivos digitais que contêm os metadados.
- USMARC (United States Machine Readable Cataloging): é um padrão para a representação e comunicação de informação bibliográfica, numa forma que possa ser lida por máquina, sendo utilizado nos Estados Unidos por praticamente todas as bibliotecas que possuam um catálogo on-line. Esse padrão sofreu várias mudanças e, fora dos Estados Unidos, foram criados alguns padrões semelhantes, tais como: o UNIMARC, o CANMARC, o UKMARC, etc..
- DIF (Directory Interchange Format): o padrão DIF, embora não seja um padrão oficial, foi, até pouco tempo atrás, um padrão de fato desenvolvido pela NASA para apoiar o Diretório Principal da NASA (NASA Master Directory) e o diretório

Principal de Mudança Global. O padrão DIF consiste em uma coleção de campos que detalham informações específicas sobre os dados, sendo compatível com o do FGDC, podendo-se, portanto, associar os elementos entre os dois padrões.

Existem outros padrões de metadados no Canadá (CGSB – Directory Information Describing Digital Geo-referenced Sets), na Europa (Method of Describing Data Sets), na Austrália e Nova Zelândia (ANZLIC – Australian and New Zealand Land Information), etc.

A complexidade dos padrões de metadados disponíveis, como os citados acima, tem frustrado muitas pessoas envolvidas com metadados. A fim de se contornar este problema, foi então sugerida a criação de um subconjunto simplificado de um padrão completo, ao qual se denomina núcleo (core), tais como os padrões Dublin Core ou Denver Core. A tendência natural é que este núcleo substitua o próprio padrão.

### **3. Descrição do Sistema Mapoteca Digital**

Os metadados têm importância fundamental em diversas situações que envolvam a transferência, documentação ou visualização de dados geoespaciais. Em particular, no contexto da Mapoteca Digital, os metadados são essenciais para auxiliar o usuário a determinar quais mapas e outros documentos estão disponíveis, como e onde encontrá-los e, principalmente, se eles servem para um determinado propósito.

O objetivo principal do Sistema Mapoteca Digital é de disponibilizar, disseminar e organizar os acervos de documentos geológicos e geofísicos da área de E&P da Petrobras, através de um sistema de banco de dados baseado em metadados. O projeto de desenvolvimento da ferramenta levou em conta os padrões de metadados existentes, principalmente o CSDGM, que se mostrou excessivamente genérico, apresentando um conjunto muito grande de informações para se descrever um dado documento.

Desse modo, optou-se por uma solução tradicional, proprietária, fundamentada no armazenamento dos metadados em um esquema de banco de dados relacional. O modelo de

dados projetado teve como principal requisito a simplicidade, de modo a facilitar a carga e manutenção dos dados.

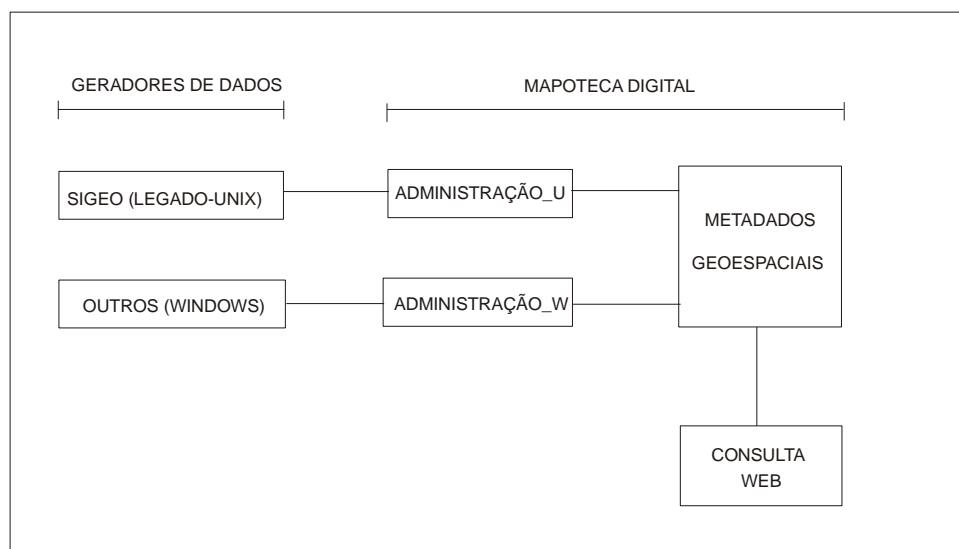


Figura 1 – Arquitetura do Sistema

### 3.1 Arquitetura do sistema

A Figura 1, acima, ilustra a arquitetura do Sistema Mapoteca Digital. Os sistemas geradores de dados espaciais são, basicamente, de dois tipos: a) Sistemas Legados, representado pelo Sistema Integrado de Geologia e Geofísica (SIGEO), implementado em plataforma Unix; b) Outros Sistemas, baseados em Windows. Para cada um destes geradores de dados, têm-se um conjunto de processos de administração, responsáveis pela inserção, alteração e exclusão dos metadados. Sobre a base de metadados, implementada em Oracle, atua um processo de consulta. Este processo, desenvolvido para Web, é responsável pela disseminação da informação.

A seguir é feita uma breve descrição da base de metadados e dos processos de administração e consulta do sistema.

### 3.2 Modelo de dados da base de metadados

A Figura 2, abaixo, descreve o modelo lógico de dados do sistema. Para cada documento catalogado, uma nova instância no banco é criada. A tabela meta\_mapoteca do modelo armazena o nome do documento, sua descrição, os limites geográficos (bounding box), escalas e responsáveis pelo o cadastramento do dado. Cada documento é classificado por um tipo, tabela tipo\_documento, que pode ser, por exemplo, um mapa de isovalores, um perfil, uma seção geológica ou sísmica, uma carta de altimetria e etc. Todos os documentos possuem uma localização que é representada por uma ou mais ocorrências nas tabelas de unidades operativas, bacias sedimentares, blocos de petróleo, países e unidades da federação. As tabelas meta\_arquivo e meta\_imagem armazenam uma imagem ou ícone do gráfico em formato jpeg. Os documentos, em sua grande maioria, encontram-se associados a um sistema de projeção cartográfica e a um referencial geodésico, representados nas tabelas meta\_projeção e datum\_geodésico, respectivamente.

Uma característica importante do sistema recai sobre o esquema de segurança implementado. Todos os metadados encontram-se disponíveis a todos os usuários, exceto a sua visualização gráfica. A tabela controle\_acesso disponibiliza a visualização de cada documento para um conjunto de usuários ou grupos de usuários.

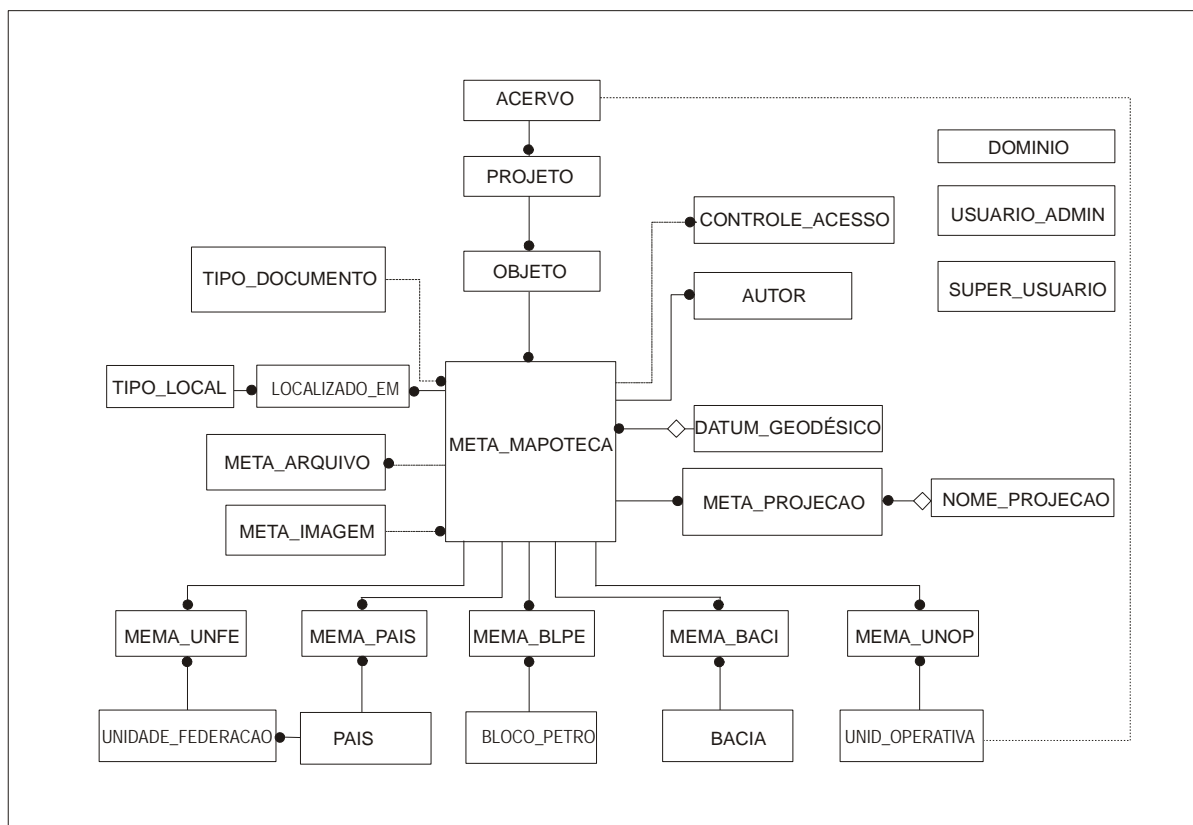


Figura 2 – Modelo de dados.

### 3.3 Processos de administração e consulta

Os metadados são alimentados através de procedimentos adaptados segundo as plataformas dos geradores de dados, Unix ou Windows. Estes processos de administração permitem que novos documentos sejam inseridos na base, excluídos ou alterados.



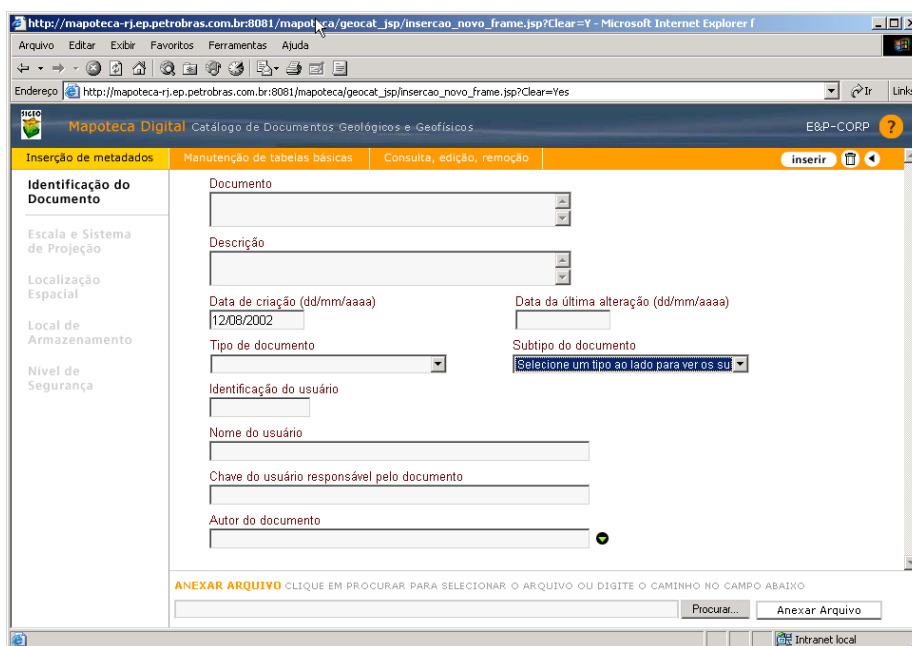


Figura 3 – Interface de administração: inserção, exclusão e alteração de metadados.

O processo de consulta atua na busca de documentos, gerando, em sua saída, coleções de itens gráficos que se assemelham aos resultados de uma consulta bibliográfica. É utilizado na busca de alguma informação existente, com o objetivo de se evitar retrabalho. Como a disseminação da informação é a característica de maior relevância, este processo foi implementado na Web, podendo ser executado via browsers. As figuras abaixo ilustram a interface com usuário do processo de consulta.

Figura 4 – Restrições de busca para consulta dos metadados: por nome do documento, descrição, autor, tipo de mapa, bacia, bloco, unidade de negócio, país e estado.

Documento	Descrição
Bloco SEAL-100 - Locacoes + Leads (antigos)-Bloco 100 - SE SES106 (Topo Res. Eoceno prof SAD69)	Bloco 100 - SE SES106 (Topo Res. Eoceno prof SAD69)
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Base Obj. Eoceno prof)	Contorno 0
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Base Obj. Eoceno prof)i	Bloco SEAL-100 - Loc.D (Base Obj. Eoceno prof)i
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Isolitas am Obj. Eoceno)i	Bloco SEAL-100 - Loc.D (Isolitas am Obj. Eoceno)i
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Isopacas Obj. Eoceno)	Bloco SEAL-100 - Loc.D (Isopacas Obj. Eoceno)
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Isopacas Obj. Eoceno)i	Bloco SEAL-100 - Loc.D (Isopacas Obj. Eoceno)i
Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Mapa Base)	mapa base de pocos, pts, xyz

Figura 5 – Resultado de uma consulta: lista de metadados.

The screenshot shows a web application window titled 'Catálogo de Documentos Geológicos e Geofísicos'. The main content area displays the following metadata for a selected document:

<b>Identificação do Documento</b>	Data de criação 20/03/2001
<b>Escala e Sistema de Projeção</b>	Documento Bloco SEAL-100 - Loc.D (play Eoceno)-Bloco SEAL-100 - Loc.D (Base Obj. Eoceno prof)
<b>Localização Espacial</b>	Descrição Contorno 0
<b>Local de Armazenamento</b>	Tipo do documento MAPA
<b>Nível de Segurança</b>	Subtipo do documento CONTORNO
	Data da última alteração 20/03/2001
	Nome do sistema gerador SIGEO

At the top of the document details, there are three buttons: 'Alterar consulta', 'Nova consulta', and 'Volta'.

Figura 6 – Metadados de um documento selecionado: identificação do documento, escalas e sistemas de projeção, localização espacial e níveis de segurança do documento.

The screenshot shows a login dialog box with the following elements:

- A title bar with the text 'Entrar'.
- Three input fields labeled 'usuário', 'senha', and 'domínio'.
- An 'OK' button at the bottom right.
- A note at the bottom: 'Obs.: Os dados acima correspondem aos da rede Windows.'

Figura 7 – Visualização da imagem condicionada ao esquema de segurança;

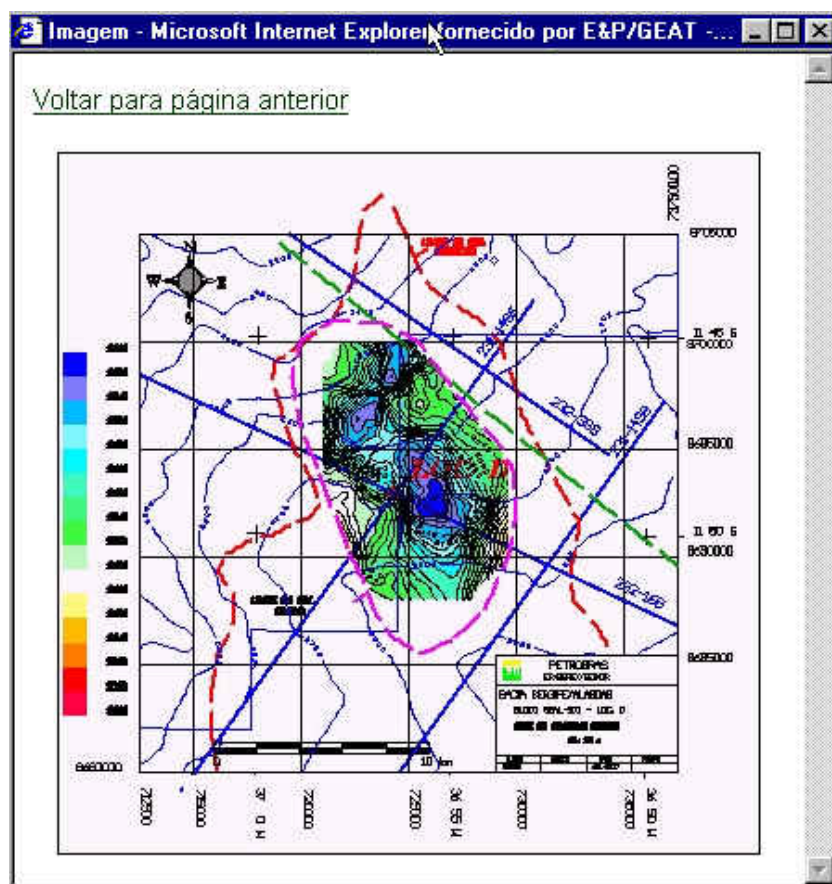


Figura 8 – Visualização da imagem armazenada: ícone.

#### 4. Conclusões

É inegável a importância dos metadados na gestão do conhecimento. Uma implementação como a que foi apresentada, baseada na Web, expande a abrangência da tecnologia, podendo, desta forma, ser classificada como intensamente relacionada com os processos de Organizar/Adquirir e Distribuir/Compartilhar, segundo a taxonomia CRIE/UFRJ.

Um fato importante, que merece ser destacado, é que as diferentes visões sobre metadados estão relacionadas, diretamente, ao estágio da organização ou instituição no contexto da gestão do conhecimento. Organizações com menor foco na gestão do conhecimento gerenciam dados

brutos, e organizações mais focadas e direcionadas para a gestão do conhecimento administram seus recursos de informação e conhecimento de forma mais profunda, controlada e organizada.

## 5. Bibliografia

ABREU, M. P., Um Estudo Classificatório das Ferramentas Tecnológicas Envolvidas em um Processo de Gestão do Conhecimento, Tese de M.Sc., Engenharia de Produção – COPPE/UFRJ, Rio de Janeiro, 2002.

ALMEIDA, L. F. B., A Metodologia de Disseminação da Informação Geográfica e os Metadados, Tese de D.Sc., Centro de Ciências Matemáticas e da Natureza – UFRJ, Rio de Janeiro, 1999.

CAVALCANTI, M., GOMES, E., PEREIRA, A., Gestão de Empresas na Sociedade do Conhecimento. Rio de Janeiro: Campus, 2001.

GUNTHER, O., Environmental Information Systems, Berlin, Springer-Verlag, 1998.

MARCO, D., Building and Managing the Metadata Repository – A Full Lifecycle Guide, New York, John Wiley & Sons, 2000.

MORIATY, T., What is Metadata? , Database Programming and Design, San Mateo, v. 10, n. 7, p.57-59, Julho, 1997.

SIMON, E., TOMASIC, A., GALHARDAS, H., A Framework for Classifying Scientific Metadata, New York, American Association for Artificial Intelligence, 1998.

TANNENBAUM, A., Metadata Solutions, New Jersey, Addison Wesley, 2002.