

Arquitetura de um banco de dados para suporte à integração de dados de campo e de sensoriamento remoto em estudos limnológicos e meteorológicos

Arley Ferreira de Souza ^{1,3}
Cláudio Clemente Barbosa ¹
Evelyn Márcia Leão de Moraes Novo ²
José Luiz Stech ²

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{arley, claudio}@dpi.inpe.br

² Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{evlyn, stech}@dsr.inpe.br

³ ETEP Faculdades, São José dos Campos - SP, Brasil
arley.souza@etep.edu.br

Abstract. The efficient use of remote sensing information must rely on ground data either to support algorithm development or for validation of the geophysical variables extracted from the images. So, most of the projects have a huge amount of ground data, both raw and processed which at end of a given research are either discarded or stored in places not accessible for other applications. All the money invested in the data collection is virtually lost, and the information available is only related to summaries, tables and conclusions about data which otherwise could be reprocessed and applied to answer many other relevant questions. Most of the time the data are stored and not shared because of the lack of a database structure, which enables its retrieval. This paper reports the steps and challenges overcome to implement a data base structure to gather limnological and meteorological data from different types and source.

Palavras-chave: remote sensing, database, spatial data model, sensoriamento remoto, banco de dados, modelo de dados geográficos.

1. Introdução

As organizações investem milhões em pesquisas que geram resultados relevantes, mas não praticam a política de garantir que os dados originais coletados, para as diversas finalidades, sejam reunidos numa base única, que permita a manutenção de sua integridade e o acesso compartilhado, por diferentes grupos de pesquisa, ao longo do tempo.

As dificuldades de reunir dados em bases unificadas começam antes mesmo da modelagem do banco, uma vez que sua criação esbarra na resistência dos pesquisadores, que colocam obstáculos ao cedê-los para o armazenamento por terceiros, embora existam protocolos que garantam que o acesso a eles será limitado às pessoas autorizadas pelo autor.

Superado o problema de reunir os dados em uma base comum, outro grande desafio é a modelagem do banco a partir de dados dispersos em planilhas e arquivos textos, organizados segundo anotações de campo e/ou a necessidade específica do responsável. Uma dificuldade particular nesse caso refere-se a dados indexados segundo toponímia local, sem anotações de coordenadas geográficas. Inúmeras vezes as anotações de localização são tão vagas que apenas quem coletou é capaz de recuperar a posição da amostra. Um exemplo dessa tendência é, por exemplo, uma descrição do tipo: “Ponto a 100m do rancho do prefeito”. Caso mais grave ainda foi relatado por Vasconcelos e Novo (2004), em que os locais anotados em planilhas de campo, em diferentes datas, deixavam de existir em decorrência do uso e ocupação da terra. Num caso como esse, a informação espacial é perdida para sempre.

A organização de dados com o atributo espacial em SIG (Sistema de Informação Geográfica) permite contextualizar uma leitura em relação ao ambiente (Câmara et al., 2000; Rigaux et al., 2002; Steinberg e Steinberg, 2006). Contudo, grande parte dos pesquisadores em campo está mais preocupada com a precisão das medidas sobre o fenômeno de interesse do que com sua posição no espaço. Atualmente, com a abordagem multidisciplinar do ambiente terrestre, essa tendência tem se modificado, talvez pela maior disponibilidade de equipamentos e softwares de localização e pela necessidade de contextualizar os dados resultantes das coletas (Stech et al., 2006). Outro fator que tem ampliado o interesse em que os dados coletados estejam associados a coordenadas geográficas é a ampliação do acesso a imagens de satélite, a qual se tornou uma ferramenta fundamental para o planejamento de missões de campo (Barbosa, 2005).

O objetivo deste trabalho, portanto, é relatar o desenvolvimento de uma base de dados limnológicos, meteorológicos e de sensoriamento remoto, que teve início para agrupar tanto dados brutos quanto processados, derivados do projeto “Balanço de Carbono nos Reservatórios de Furnas Centrais Elétricas S.A.”¹. Atualmente, essa base se expandiu para incluir dados do projeto GEOMA (Rede Temática de Pesquisa em Modelagem Ambiental da Amazonia) e dados coletados para dissertações e teses. Uma das vantagens de uma base unificada é permitir que se façam consultas que integrem os diferentes tipos de dados armazenados. Essa unificação também torna mais eficiente a modelagem de uma interface única de acesso remoto. Não bastando apenas uma estrutura para armazenar os dados, mas também a necessidade de um sistema que permita sua disponibilidade para os usuários.

O artigo descreve o processo de modelagem do banco de dados e da interface de disponibilização na Internet, na qual se permitem fazer consultas personalizadas e tanto visualizar os resultados quanto exportá-los nos formatos: texto, planilha e KML (Keyhole Markup Language).

2. Natureza dos Dados

Os dados armazenados são resultados de coletas de campo de parâmetros limnológicos e meteorológicos, tais como, temperatura da água e do ar, intensidade do vento, pH e concentração de oxigênio dissolvido na água.

As medidas em campo se caracterizam pelo mesmo tipo de parâmetro, por exemplo, temperatura da água, ser coletado em diversos locais e, em alguns casos, em diferentes profundidades. A profundidade das leituras pode variar de um ponto de coleta para outro em função do nível da água ou algum parâmetro ambiental, assim como a transparência da água (Kirk, 1994; Novo et al., 2006). Além disso, os mesmos parâmetros podem ser coletados, na mesma posição espacial, em diferentes épocas, com a finalidade de se compor séries temporais.

Devido a possíveis correlações entre parâmetros distintos, o que muitas vezes ajuda a explicar os processos em estudo, algumas coletas incluem a leitura de muitos parâmetros num mesmo espaço-tempo.

Os valores dos parâmetros são, em sua maioria, numéricos e, em alguns casos, incluem um texto descrevendo o contexto da leitura. Podendo existir arquivos associados às coletas, tais como, fotografias, imagens de satélite, arquivos bruto dos dados coletados (no formato gerado pelo equipamento de coleta).

A forma de coleta dos dados pode ocorrer de duas maneiras: através de trabalhos de campo, onde pessoas medem alguns parâmetros por um curto período de tempo e os registram em planilhas e arquivos textos; e a outra é através de equipamentos instalados em campo, que fazem a coleta periódica e automática dos dados e os transmitem via satélite. Atualmente, no

¹ <http://www.dsr.inpe.br/projetofurnas/>

último formato, são recebidos os dados do SIMA (Sistema Integrado de Monitoramento Ambiental) que são parâmetros limnológicos e meteorológicos coletados em lagos naturais e artificiais (Stech et al., 2006).

3. Modelo de Dados

Uma característica comum de todo dado geográfico é possuir o atributo posição geográfica na superfície da terra. Esta característica foi o ponto de partida para modelar uma base de dados (Figura 1), cujo objetivo seria armazenar dados oriundos de diferentes fontes, coletados em diferentes formatos.

Os pesquisadores, entretanto, por motivo de facilidade, sempre associam a posição geográfica a um nome de fácil memorização. No modelo aqui proposto, este nome foi denominado de “sítio”. A partir da definição de sítio, foi possível criar uma estrutura de banco onde todas as tabelas, que possuem dados ambientais, relacionam-se com a tabela *Sitio*, tabela esta que contém a identificação dos locais de coleta (nome, longitude, latitude e descrição do local).

Algumas vezes, a identificação do sítio é associada, pelo pesquisador, ao nome da região a qual ele pertence, como por exemplo, “Represa de Serra da Mesa” e “Lago de Curuai”. Como em um local ou região existem vários sítios, então foi criada a tabela *_Local* para armazenar o nome destas regiões de coleta. Assim como o sítio, a região, aqui denominada de local, também pode ter um texto descritivo associado (representado no modelo da Figura 1 pelo campo *descricao*).

Normalmente, cada grupo de dado, possui atributos específicos. Por exemplo, o grupo “gases” pode possuir atributos diferentes do grupo “parâmetros limnológicos”, que por sua vez pode ser diferente do grupo “dados meteorológicos”. Para que a base de dados possa ter a flexibilidade necessária de cada grupo, o sistema em desenvolvimento permite criar uma tabela para cada grupo, como ilustrado para o caso de *Gases* e *Parlimnologicos* na Figura 1. As tabelas do tipo grupo apresentadas na Figura 1 são apenas ilustrativas e não apresentam todos os campos relativos aos parâmetros coletados.

Tipos de dados que possuem cota de medida (medida em altura/profundidade) como atributos são representados como um campo-chave da tabela, isto permite separar leituras distintas no mesmo espaço-tempo.

Para permitir que a base de dados possa tratar o conceito de campanha (ida esporádica a um determinado local/região para realizar medidas), foi criada a tabela *Campanha* (Figura 1) para armazenar o período e o número da campanha em um determinado local. O número da campanha é pertinente porque o usuário, às vezes, necessita resgatar os dados por campanha.

Para que o sistema gerenciador de banco de dados possa tratar de forma eficiente, dados provenientes de equipamentos que geram séries temporais, como o caso do SIMA, mencionado na seção anterior, ou seja, um grande conjunto de medidas em um mesmo sítio, criou-se o conceito de tabela *xxx_leituras* (onde *xxx* é o nome do equipamento que gera a série temporal) para armazenar as medidas automáticas e a tabela *xxx* para armazenar os campos que não se alteram de uma medida para outra.

Esta solução evita a redundância de dados e informações. Como no caso das outras tabelas citadas acima, a tabela *xxx* se relaciona com a tabela *Sitio*, que conterá a posição geográfica da série temporal. Para contemplar o caso de sistemas automático com links por satélites, a tabela *xxx* possui o campo que registra o número da *pcd* (transmissor instalado na estação de coleta automática).

O banco possui ainda uma área para o catálogo de imagens, mas que atualmente esta sendo utilizada apenas para o armazenamento de imagens ALOS/PALSAR². A tabela *Imagem*

² <http://www.eorc.jaxa.jp/ALOS/about/palsar.htm>

possui o campo *localArmazenamento* para armazenar a URL (Uniform Resource Locator) de onde a imagem se encontra.

Os dados foram organizados em dois bancos, em um deles estão os dados coletados (Figura 1) e no outro estão as informações que são utilizadas no ambiente de acesso aos dados (Figura 2).

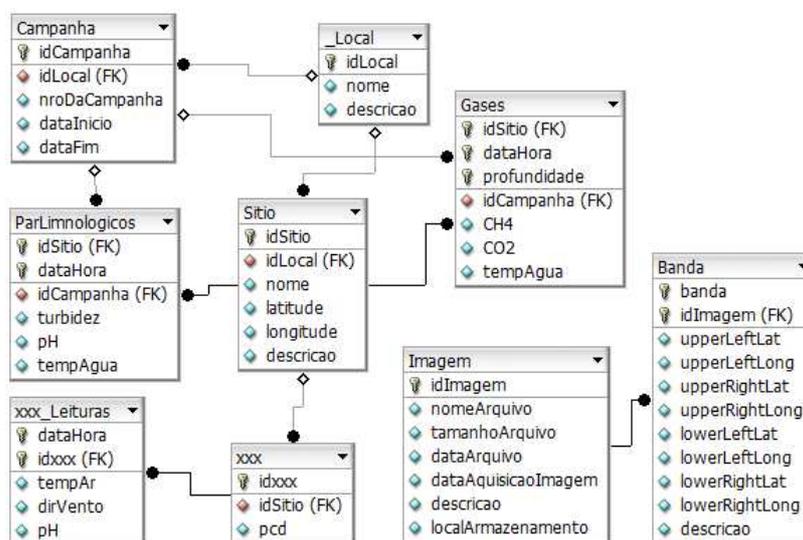


Figura 1. Modelo de banco para o armazenamento dos dados coletados.

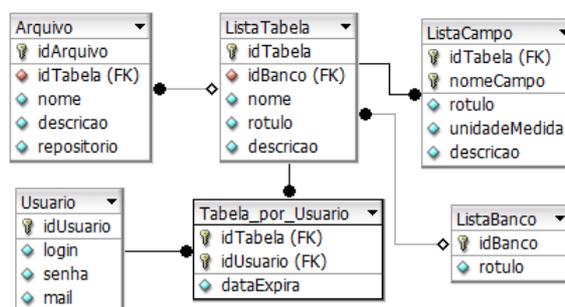


Figura 2. Modelo de banco para o armazenamento do controle de acesso e estrutura dos demais dados armazenados.

Na Figura 2, a tabela *ListaTabela* armazena a lista de tabelas que existe no banco, onde uma das suas utilidades é conter o rótulo que terá a tabela ao ser apresentada para o usuário no ambiente de acesso. Por exemplo, a tabela de nome *ParLimnologicos* poderá ser apresentada para o usuário com o rótulo “Parâmetros limnológicos” ou “Limnological parameters”, tudo depende do rótulo cadastrado. O campo *descricao* é usado para armazenar algum texto explicativo sobre a tabela.

A tabela *ListaCampo* armazena a lista de campos que existe em cada tabela, o objetivo desta tabela é dar rótulo, unidade de medida e um texto descritivo a cada campo da tabela ao exibi-lo no ambiente de acesso.

A Figura 3 apresenta a tela do ambiente de acesso onde o usuário escolhe qual banco e tabela ele quer acessar. Para formar o conceito de banco foi criada a tabela *ListaBanco* que armazena o rótulo dos supostos bancos que agrupam as tabelas. Esta estrutura além de tornar mais fácil o acesso a uma determinada tabela, ela faz com que o usuário tenha a impressão que seus dados estão em um banco separado.

A lista de usuários foi armazenada na tabela *Usuario* e os acessos dos usuários na tabela *Tabela_por_Usuario*. O campo *dataExpira* é utilizado para limitar o período de acesso do usuário a uma tabela. Da forma como está estruturado o banco, o controle de acesso é limitado por tabela.

A tabela *Arquivo* armazena arquivos e os relaciona com alguma tabela, isto é útil porque o usuário pode querer incluir programas de manipulação dos dados, imagens etc. juntamente com os dados alfanuméricos.



Figura 3. Tela para selecionar o Banco e a Tabela no ambiente de acesso.

4. Ambiente de Acesso aos Dados

O ambiente de acesso aos dados foi implementado utilizando a linguagem PHP e o Sistema Gerenciador de Banco de Dados foi o MySQL 5.

Ao acessar o portal é exibida uma tela para o usuário se conectar. Os acessos que serão disponibilizados para o usuário são aqueles cadastrados na tabela *Tabela_por_Usuario*. A Figura 3 exibe o “menu” de acesso aos bancos e tabelas de um usuário cadastrado. Ao acessar uma tabela, o “menu” desta tabela será exibido, veja um exemplo na parte esquerda da Figura 4.

Parâmetros do Banco

Banco: Geoma

Tabela: ADCP

[Acessar](#) [Configurações](#)

Acesso a ADCP

Local: Curuai

Campanhas: Segunda

Visualizar os nomes dos Sítios

Exibir coordenadas em:

GMS GD Nenhuma

Busca usando coordenadas

Busca usando identificação de local

Local:

Busca em um faixa de horário

Busca em um intervalo de dias

Busca pelo valor de um sensor

Sensor: Depth1

valores entre 0.8 e 0.9

Incluir campos sem valor

[Tabela](#) [Arquivos](#)

Dados de ADCP da 2ª campanha em Curuai

Listando as leituras de "Depth1" que possuem valores entre 0.8 e 0.9

Nro	Profile	Data	Horário	Distance (m)	DMG (m)	Dist. east (m)	Dist. north (m)	Latitude	Longitude	Depth1 (m)	Depth2 (m)
1	4	2006-09-13	15:04:11	The distance made good (i.e., straight-line distance) from the data set starting point (0,0 coordinates) to this profile							
2	5	2006-09-13	15:04:56	5.9	4.2	-1.5	3.9	-2.24274	-55.0666	0.9	0.99
3	5	2006-09-13	15:51:05	8.1	6.1	0.5	-6.1	-2.23692	-55.0784	0.85	1.22
4	6	2006-09-13	16:16:40	6.1	6.1	-5.8	1.9	-2.23477	-55.0707	0.86	0.9
5	1	2006-09-13	16:01:19	0.2	0.2	-0.2	0.2	-2.2407	-55.0787	0.87	0.8
6	2	2006-09-13	16:01:25	1	1	-0.6	0.8	-2.2407	-55.0787	0.89	0.81
7	1	2006-09-13	16:12:35	1.3	1.3	-0.2	1.3	-2.23506	-55.0708	0.9	1.25
8	2	2006-09-13	16:12:41	2.8	2.8	-0.6	2.8	-2.23506	-55.0708	0.87	1.23

O DATUM é WGS84.
O horário das leituras é UTC.

Total de registros desta tabela: 8 (1 página)
Página atual: 1

[Início](#) [Anterior](#) [Próxima](#) [Final](#)

Ir para a página [ok](#)

Número de linhas da tabela 60 [Mostrar a tabela](#)

[Google Earth \(.kml\)](#) [Excel 2003 e XP](#) [Exportar dados \(.txt\)](#)

Figura 4. Tela de acesso aos dados.

Como a maioria das tabelas possui os campos: local, campanha, data e horário; foi possível padronizar uma tela para o usuário personalizar consultas aos dados. A opção *Sensor* (na parte inferior esquerda da Figura 4) lista os campos de dados que possuem a tabela em questão (neste exemplo, a tabela ADCP), estes campos foram obtidos na tabela *ListaCampo*.

Ao clicar no botão “Tabela” a tabela da parte direita da Figura 4 será exibida. Nela são listados os registros que satisfazem os parâmetros da consulta. No cabeçalho desta tabela aparecem os rótulos e as unidades de medida de cada coluna. Ao posicionar o cursor sobre o cabeçalho é exibida a descrição da coluna, veja como exemplo, o texto associado à coluna DMG que é “The distance made...”. Estes valores foram obtidos nos campos *rotulo*, *unidadeMedida* e *descricao* da tabela *ListaCampo*. Na parte inferior desta mesma tabela aparece o texto “O DATUM é...”, este texto foi obtido no campo *descricao* da tabela *ListaTabela*. Os dados exibidos na tabela podem ser exportados como arquivo texto, planilha e a localização dos pontos podem ser exportadas para o Google Earth.

A Figura 5 lista na parte direita os arquivos que foram associados à tabela ADCP. Estas informações foram obtidas na tabela *Arquivo*.

Parâmetros do Banco

Banco: Geoma
Tabela: ADCP
Acessar Configurações

Acesso a ADCP

Local: Curuai
Campanhas: Segunda

Visualizar os nomes dos Sítios

Exibir coordenadas em:
 GMS GD Nenhuma

Busca usando coordenadas
 Busca usando identificação de local

Local:

Busca em um faixa de horário
 Busca em um intervalo de dias
 Busca pelo valor de um sensor

Sensor: Depth1
valores entre 0.8 e 0.9
 Incluir campos sem valor

Tabela Arquivos

Documentos relacionados com a tabela 'ADCP'

Nome do arquivo	Descrição do arquivo
RiverSurveyor.pdf	Manual do SonTek, versão 4.3
RiverSurveyour_Setup.zip	Software para manipular medidas do ADCP. Inclui o setup de

Figura 5. Tela de acesso aos arquivos.

A interface de acesso às imagens é exibida na Figura 6. A tela de consulta permite consultar esta base de dados e selecionar as imagens de interesse por data, polarização e coordenada de interesse central. A cobertura das imagens selecionadas pode ser exportada para o Google Earth (formato KML).

Parâmetros do Banco

Banco: Files
Tabela: Imagens
Acessar Configurações

Acesso to Imagens

Search using part of filename:

Search using a point:
Lat.: -2.08
Long.: -55.10

Show box coordinates in:
 Degree DMS

Search using date interval:
From: 1 1 2008
To: 31 12 2008
 Include files without date

Polarization: HH

Dados de Imagens

Images between 2008-01-01 and 2008-12-31
Polarization HH
Images next to (-2.08, -55.10)

Nro	Filename	Date	Polarization	Upper left lat	Upper left long	Upper right lat
1	KC_003-16406N08S25WB1SLT1.tar.gz	2008-01-13	HH	6.78427219	-50.33219910	7.4696141
2	KC_003-16409N08S25WB1SLT1.tar.gz	2008-01-18	HH	7.14598036	-51.86618805	7.828663
3	KC_003-19406N08S16WB1SLT1.tar.gz	2008-05-30	HH	6.76626779	-50.34286118	7.435774
4	KC_003-19409N08S16WB1SLT1.tar.gz	2008-06-04	HH	7.14785385	-51.86725616	7.817874
5	KC_003-20409N09S17WB1SLT1.tar.gz	2008-07-20	HH	7.57900953	-51.77285385	8.251596
6	KC_003-21406N09S21WB1SLT1.tar.gz	2008-08-30	HH	7.74315119	-50.13478851	8.418361
7	KC_003-21409N08S16WB1SLT1.tar.gz	2008-09-04	HH	7.20635462	-51.85453415	7.879265
8	KC_003-22406N08S30WB1SLT1.tar.gz	2008-10-15	HH	6.79147148	-50.32365036	7.482315

Total de registros desta tabela: 8 (1 página)
Página atual: 1

Início Anterior Próxima Final

Ir para a página sk

Número de linhas da tabela 60 Mostrar a tabela

(Google Earth (.kml)) Excel 2003 e XP Exportar dados (.txt)

Figura 6. Tela de consulta de imagens e visualização do recobrimento das cenas selecionadas no Google Earth.

5. Conclusões

O trabalho mostrou um modelo de dados que foi desenvolvido para o armazenamento e disponibilização de dados ambientais.

A arquitetura de banco de dados desenvolvida neste trabalho pode ser utilizada para outras áreas do conhecimento, nos formatos tabulares, binários e matriciais (imagens).

Este banco de dados permite análises rápidas das variáveis armazenadas, por meio de consultas estruturadas permitindo planejamentos futuros.

Referências Bibliográficas

- Barbosa, C. C. F. **Sensoriamento remoto da dinâmica de circulação da água do sistema planície de Curai/Rio Amazonas**. 2005-12-09. 286 p. (INPE-14614-TDI/1193). Tese de Doutorado - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2005. Disponível em: <<http://urlib.net/sid.inpe.br/MTC-m13@80/2006/02.22.15.03>>. Acesso em: 06 nov. 2008.
- Câmara, G. **Anatomia de um SIG**. Fator GIS, ano 1, n. 04, p. 11-15, 1994.
- Câmara, G.; Souza, R. C. M.; Pedrosa, B. M.; Vinhas, L.; Monteiro, A. M. V.; Paiva, J. A.; Carvalho, T. M.; Gattass, M. **Terralib: Technology in support of GIS innovation**, II Workshop Brasileiro de Geoinformática, Geoinfo, Caxambu, 2000.
- Kirk, J.T.O. **Light and photosynthesis in aquatic ecosystems**. 2^oed. Cambridge: University Press, 1994.
- Novo, E.L.M.M.; Barbosa, C.C.F.; Freitas, R.M.; Shimabukuro, Y.E.; Melack, J.M.; Pereira-Filho, W. **Seasonal changes in chlorophyll distribution in Amazon floodplain lakes derived from MODIS images**. Jap. J. Limn., 7, 153-161, 2006.
- Rigaux, P.; Scholl, M.; Voisard, A. **Spatial databases with applications to GIS**. Morgan Kaufmann, 2002.
- Stech, J. L.; Lima, I. B. T.; Novo, E. M. L. M.; Silva, C. M.; Assireu, A. T.; Lorenzetti, J. A.; Carvalho, J. C.; Barbosa, C. C. F.; Rosa, R. R. **Telemetric monitoring system for meteorological and limnological data acquisition**. Verh. Internat. Verein. Limnol. 29, 1747-1750, 2006.
- Steinberg, S. J. e Steinberg, S. L. **GIS: Geographic Information Systems for social sciences: investigating space and place**. SAGE, 2006.
- Vasconcelos, C. H.; Novo, E. M. L. M. **Mapeamento do uso e cobertura da terra a partir da segmentação e classificação de imagens fração solo, sombra e vegetação derivadas do modelo linear de mistura aplicado a dados do sensor TM/Landsat5, na região do reservatório de Tucuruí - PA**. Acta Amazonica, Manaus, v. 34, p. 487-493, 2004.