

## **Estudo comparativo entre diferentes modelos de organização de acervos de dados eletrônicos**

**Rinalda Francesca Riecken**

***Doutoranda em Ciência da Informação pela Universidade de Brasília; Mestre em Gestão do Conhecimento e Tecnologia da Informação pela Universidade Católica de Brasília (2002); Consultora no Ministério da Previdência Social para assuntos de tecnologia de informação (2003-).***

*Realiza um estudo comparativo entre diferentes modelos de organização de acervos eletrônicos. A pesquisa identificou três tipos: desconcentrado, concentrado e híbrido. A sua adoção é dependente de contexto e de fatores estratégicos. O Modelo Desconcentrado apresenta-se como o mais adequado na relação custo-benefício; o Modelo Concentrado, o de melhor governança, embora utilize, geralmente, tecnologias com padrões não abertos; e o Modelo Híbrido, em longo prazo, o mais flexível e menos vulnerável frente a evoluções tecnológicas, contudo, o mais dispendioso.*

***Palavras-chave: Ciência da Informação (CI); Organização da Informação; Modelo de Informação, Arquitetura de Informação; Arquitetura de Tecnologia de Informação (ATI), Gestão de Conteúdo.***

### **Comparative study between different organization models of large electronic data**

*A comparative study between different models for organization of large electronic collections is shown. The research identified three types of organization: decentralized, centralized and hybrid. The adoption of each one depend on context and strategic factors. The decentralized model is more appropriate due to the better*

*cost-benefit relationship; the concentrated model offers more governance, but it uses proprietary standard technologies; and the hybrid model in the long term is the most flexible and least vulnerable to changing technologies, although it is the most expensive.*

**Keywords:** *Information science (IS); Information organization; Information models; Information architecture (IA); Information technology architecture (ITA), Content Management (CM).*

**Recebido em 30.12.2007 Aceito em 18.06.2008**

## **Introdução**

A expansão do uso das tecnologias de informação e comunicação permitiu a acumulação de grandes volumes de informação em meio eletrônico, denominados bases ou acervos eletrônicos. A maneira como essas bases são organizadas (arquitetura informacional) e a tecnologia empregada (arquitetura tecnológica) são temas em aberto na Ciência da Informação. Assim, esse trabalho pretende contribuir para a literatura, apresentando uma taxonomia para a classificação dessas arquiteturas.

Nas situações envolvendo grandes bases ou acervos eletrônicos, os modelos de arquitetura informacional e de arquitetura tecnológica aumentam em complexidade, em função: das preocupações em atender às necessidades dos usuários, quanto aos conteúdos (gestão de conteúdos), e dos patrocinadores (gestão do negócio); da observância de padrões que permitam a adequada recuperação e troca das informações (interoperabilidade); da necessidade de medir o esforço e os custos para a operacionalização (condições e meios para a operacionalização); e do atendimento aos requisitos de desempenho (tempo de resposta, usos concorrentes, etc.), além de outros aspectos.

Por um lado, a arquitetura informacional determina os aspectos da operacionalização, por outro lado, ela também é afetada pela arquitetura de Tecnologia de Informação e Comunicação (TIC), sendo esses dois elementos, portanto, indissociáveis.

O trabalho está concentrado na identificação de parâmetros que definam uma taxonomia, bem como na análise de bases de dados, prospectadas na *Internet*, que possuam características comuns nos modelos utilizados para a organização de grandes acervos eletrônicos.

## **Modelo empírico**

A partir da literatura, dos estudos de LIMA-MARQUES (2000) e da experiência da autora em consultoria na área de TI, verificou-se que as variáveis utilizadas para a compreensão das arquiteturas de informação e tecnológica são:

- Finalidade – diz respeito à finalidade das bases de dados, no suporte aos processos de trabalho, e à decisão sobre a estrutura e gerenciamento das funções sistêmicas;
- Usuários – são as pessoas que utilizam as bases de dados eletrônicas;
- Escopo da Informação – refere-se à delimitação clara dos componentes, dos requisitos, da expectativa dos gestores/patrocinadores, e das necessidades de informações para o negócio;
- Estratégia de TIC – diz respeito às escolhas tecnológicas vis-à-vis às expectativas dos gestores/patrocinadores quanto ao prazo para disponibilizar as bases;
- Tecnologias – as tecnologias de informação e comunicação são o *hardware*, o *software*, a infra-estrutura de rede e os bancos de dados com os conteúdos armazenados; e
- Esforço para Construção e Operacionalização – refere-se ao esforço interno e de articulação externa para o desenvolvimento do sistema e a implantação dos conteúdos nas bases.

A partir da combinação dessas variáveis, identificam-se três modelos possíveis: um Modelo Desconcentrado, um Concentrado e um Modelo Híbrido, no aspecto da gestão dos conteúdos e do negócio (informacional) e no aspecto da gestão das tecnologias (operacionalização).

## **Modelo desconcentrado**

O modelo de arquitetura informacional Desconcentrada pode ser descrito sob os seguintes aspectos:

- Finalidade – presume um processo de interação entre as bases de dados geograficamente dispersas, propiciando uma coerência corporativa que, no caso das diversas esferas de governo, oferece um arranjo colaborativo e de suporte padronizado ao sistema visando ao acesso a bases de dados remotas;
- Usuários – o impacto da expressiva quantidade de usuários que buscam o mesmo serviço informacional simultaneamente (uso concorrente) é minimizado devido à distribuição geográfica das bases;
- Escopo da Informação – é amplo, abrangendo não só os registros de interesse do patrocinador, mas também os conteúdos específicos desejados pelos cooperados;
- Estratégia de TIC – adota uma proposta colaborativa (ambiental), apresentando soluções já largamente utilizadas, de baixo custo e favoráveis a longo prazo;

- Tecnologias – são padronizadas pelos patrocinadores/gestores na forma de um *menu* de opções limitadas, que podem ser selecionadas pelos cooperados que aderem ao modelo; e
- Esforço para Construção e Operacionalização – sistema pouco complexo devido ao uso de padrões de consenso entre as partes, podendo envolver acordos formais para o estabelecimento de padrões de documento e tecnológicos. A articulação pode se tornar difícil na medida em que as soluções genéricas não colhem especificidades e desejos de personalização, comuns no desenvolvimento de sistemas de informações, decorrentes em geral da cultura organizacional e discricionária. A operacionalização fica minimizada e distribuída entre os gestores que aderem ao processo desconcentrado.

De um modo geral, o desenvolvimento de soluções desconcentradas é de baixa complexidade para os cooperados, com menor custo na operacionalização da infra-estrutura de TIC e na manutenção da equipe de suporte. Contudo, com maior complexidade na gestão das bases distribuídas e menor qualidade nas pesquisas, em virtude de se tratar de aplicações genéricas para atender a um grande conjunto de organizações.

## **Modelo concentrado**

A arquitetura informacional Concentrada pode ser descrita sob os seguintes aspectos:

- Finalidade – suporte a processos internos de trabalho envolvidos; as funcionalidades são normalmente estabelecidas pelo patrocinador; suporte a processos de trabalho departamentais, WEB portais corporativos de comércio eletrônico (*e-commerce*), serviços à clientela interna; referência para consultas;
- Usuários – serviços congregados em portais na *Internet*, servindo de referência para consultas públicas por usuários que não cooperam com o processo; há concentração de acesso por uma grande quantidade de usuários a um mesmo local, podendo gerar gargalos;
- Escopo da Informação – estreito, ficando restrito aos temas específicos do negócio;
- Estratégia de TIC – focada nas atividades burocráticas da organização patrocinadora para atender a seus projetos de curto prazo; torna-se desfavorável no longo prazo, devido ao uso de tecnologias proprietárias e de acervos, muitas vezes na forma de assinatura;
- Tecnologias – em geral proprietárias, especializadas e de custo expressivo; entretanto, com possibilidade de melhor usabilidade e mecanismos de busca devido ao seu uso por inúmeros clientes, sendo tecnologias comercializadas em pacotes ou assinaturas; e

- Esforço para Construção e Operacionalização – sistema mais simples por não envolver a articulação com parceiros; possui um caráter doméstico (intramuros) e um esforço concentrado, interno ou contratado, podendo ter soluções prontas contratadas no mercado, para sua operacionalização; torna-se mais simples por não envolver articulação com parceiros.

De um modo geral, o Modelo Concentrado apresenta-se como uma solução menos complexa, com custo de infra-estrutura de TIC maior, bem como com custo da equipe de suporte igualmente maior, devido à centralização da solução. Entretanto, o processo de governança em informações e nas tecnologias torna-se mais simples, bem como a qualidade da solução tende a atender melhor ao usuário patrocinador e aos clientes, quando os níveis de serviço contratados são adequadamente estabelecidos e monitorados.

Diferente da Arquitetura Desconcentrada, voltada para a cooperação, na Arquitetura Concentrada o patrocinador responsabiliza-se pela definição do modelo de informação e de TIC, seu desenvolvimento e seus custos, sem preocupação em atender a características locais. Esse modelo é, normalmente, herança de estruturas organizacionais verticais e estabelecidas há décadas, com expressivo volume de bases legadas.

## **Modelo de arquitetura híbrida**

A arquitetura informacional Híbrida pode ser descrita sob os seguintes aspectos:

- Finalidade – reúne características das duas arquiteturas anteriores, capitalizando as vantagens e reduzindo as desvantagens, no suporte aos processos internos de trabalho e de integração externa;
- Usuários – serviços podem ser cooperados utilizando-se de arquitetura flexível com bases dispersas congregadas em portais na *Internet*; possibilita um arranjo transparente para os usuários com recursos de pesquisa que localizam as informações onde quer que as bases estejam;
- Escopo da Informação – amplo, conforme as necessidades, oriundo de bases desconcentradas e geograficamente dispersas, integradas pela camada intermediária;
- Estratégia de TIC – adota uma proposta híbrida, podendo conviver em bases de tecnologias proprietárias – por exemplo, bancos de dados textuais – e não proprietárias, o que torna o processo mais favorável no longo prazo;
- Tecnologias – de plataforma aberta; possuem componentes que encapsulam programas executáveis de tecnologias proprietárias ou legadas, tornando-se transparente para os usuários as soluções tecnológicas de cada aplicativo. A Arquitetura

Tecnológica é implementada em camadas: camada de aplicações, camada intermediária (*Middleware*), camada dos dados (várias mídias); e

- Esforço para Construção e Operacionalização – por congregarem legados de várias tecnologias, há necessidade de desenvolvimento da arquitetura integradora intermediária e da camada de aplicação para comunicação das recuperações, independentemente de onde as bases de dados se encontrem; isso envolve negociação com parceiros e colaboradores, investimentos do patrocinador e uma agenda de atividades colaborativas envolvendo acordos formais e gerenciamento articulado do patrocinador da integração.

De um modo geral, o Modelo Híbrido é o mais complexo, requerendo investimentos em infra-estrutura de TIC para estruturação dos componentes integradores. Entretanto, uma vez estruturados, requer menor custo de suporte devido ao fato das bases de dados estarem distribuídas e terem gestores particulares, tornando, também, a gestão local menos complexa. Quanto à qualidade das pesquisas, depende do desenvolvimento e da disponibilização dos componentes pelos gestores que cooperam com as tecnologias.

## Metodologia

O trabalho está concentrado em estudos de algumas bases de dados identificadas, de expressivas dimensões e concorrência de usuários na recuperação das informações. A síntese desses estudos permitiu construir um quadro comparativo objetivando auxiliar, eventualmente, na escolha da arquitetura adequada para grandes acervos eletrônicos.

A prospecção realizada na *Internet* apontou características comuns dos modelos utilizados para a organização de grandes acervos eletrônicos, sendo cunhados, para fins do presente trabalho, Modelo Descentralizado, Modelo Centralizado e Modelo Híbrido.

Os estudos concentraram-se em acervos textuais normativos e afins, armazenados em formato eletrônico, comuns na administração pública e em bases de dados públicas, envolvendo expressivo volume de informações textuais, dentre outras, e de usuários. As conclusões obtidas aplicam-se, certamente, a acervos de outra natureza, por analogia.

A pesquisa permitiu identificar as características dos modelos mais utilizados, alguns pontos críticos observados e a adesão ou não aos padrões abertos e portáteis que possibilitam a eventual interoperabilidade entre sistemas/dados. São apresentados alguns exemplos práticos visando ilustrar as características dos modelos identificados.

As vantagens e desvantagens na adoção dos modelos são identificadas, utilizando-se do modelo paramétrico simplificado, sem esgotar o assunto, a fim de auxiliar na escolha do modelo de arquitetura de informação e de TIC; momento este em que, naturalmente, outras variáveis precisam ser consideradas, em especial quanto aos aspectos

estratégicos e políticos, às diretrizes ligadas às informações envolvidas e às recomendações tecnológicas em cada caso.

## **Análise empírica**

Utilizando-se os critérios identificados e a taxonomia proposta, foram analisados alguns exemplos de bases de dados eletrônicas, conforme exposto a seguir.

## **Modelos desconcentrados**

O Modelo Desconcentrado utiliza-se normalmente de um portal com um arranjo de bases e serviços, formando um conglomerado disperso geograficamente; o que resulta em um modelo cooperativo em que cada membro colabora para o resultado do acervo disponibilizado qualitativa e quantitativamente, bem como se beneficia de seus resultados.

Exemplo desse tipo é o modelo fomentado pelo Centro Especializado da Organização Pan-Americana de Saúde – OPAS (BIREME, 2006), que promove a padronização e o estímulo à estruturação de bibliotecas virtuais e aos padrões informacionais técnico-científicos em saúde, tendo sua origem na Biblioteca de Medicina, em São Paulo. Exemplificando, podem-se conhecer a legislação produzida dos órgãos e das entidades que utilizam o padrão BIREME<sup>1</sup>, suas coleções de publicações monográficas e periódicas, folhetos, cartazes, vídeos e os documentos técnico-científicos relacionados às atividades de saúde nos países envolvidos, recorrendo ao sítio da BIREME.

Ao ingressar na Biblioteca Virtual em Saúde (BVS), os usuários podem navegar no espaço das várias fontes de informação geradas, atualizadas e armazenadas na *Internet*, obedecendo a metodologias comuns que integram uma rede de entidades que se utilizam do modelo. A usabilidade, nesse modelo, atende às necessidades básicas. No aspecto das tecnologias, certa perda de desempenho pode ser observada devido à obrigatoriedade da utilização de tecnologias específicas (servidor *WEB* e banco de publicação).

Alternativamente, as bases dos colaboradores podem ser selecionadas diretamente, como, por exemplo, a BVS do Ministério da Saúde – MS (BVS/MS, 2006). O acervo da BVS/MS contém os atos publicados pelo MS, de 1998 a 2003, mas a legislação e normas brasileiras na área encontram-se dispersas nas três esferas de governo, necessitando serem solicitadas diretamente às fontes, o que pode se tornar uma tarefa difícil. Componentes especializados em saúde podem, ainda, divulgar textos científicos por meio da metodologia Literatura Latino-Americana e do Caribe em Ciências da Saúde - LILACS, um

---

<sup>1</sup> A tecnologia disponibilizada para as bibliotecas virtuais em saúde, pela BIREME, é o WWWIsis (WWWIsis, 2006), servidor de Internet que utiliza uma linguagem baseada em XML (eXtensible Markup Language), com capacidade de processamento por meio da transferência de dados de um navegador WWW (World Wide WEB).

conjunto de descritores utilizados pela BIREME que propicia a criação de bases de dados locais e nacionais (LILACS, 2006).

Outro exemplo é a *Scientific Electronic Library Online* (SciELO), uma biblioteca eletrônica de periódicos científicos brasileiros, resultado da iniciativa da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) em parceria com a BIREME e com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Tem por objetivo o desenvolvimento de uma metodologia comum para a preparação, armazenamento, disseminação e avaliação da produção científica em formato eletrônico, por meio do acesso a textos completos de artigos científicos produzidos pelos países da América Latina e Caribe, oferecendo a possibilidade de visualização do texto na íntegra (SciELO, 2006).

A Agência Nacional de Vigilância Sanitária (ANVISA) aderiu à metodologia de bibliotecas virtuais em saúde através do seu sistema VISALEGIS, em parceria com a BIREME. A base de dados é atualizada diariamente, oferecendo legislação consolidada relacionada à vigilância sanitária em âmbito federal, estadual e municipal. A alimentação é feita de forma cooperada com as Vigilâncias Sanitárias Estaduais e Municipais. O Sistema disponibiliza uma base de dados com textos completos para pesquisa (ANVISA, 2006).

O QUAD 1, a seguir, ilustra as características do Modelo Desconcentrado. Podem-se resumir os pontos importantes desse Modelo:

- O Modelo BIREME e suas variantes constituem alternativa simples e barata, dentro de um processo cooperativo, onde cada órgão/entidade que adere ao modelo gerencia sua base e pode desenvolver *interfaces* nos moldes de um hipertexto.
- O principal aspecto desse modelo é que não há necessidade de se ter todos os bancos de dados em um único servidor, cuja escalabilidade seria um grande problema.
- O importante é a divulgação dos *links*, nos moldes dos *sites* de busca, e a conectividade entre as bases de informações.
- É um modelo interessante para processos cooperativos onde os atores não dispõem de recursos tecnológicos expressivos, e verifica-se um processo natural de "capilaridade" na disseminação dos padrões informacionais e tecnológicos entre as diversas esferas de governo e no âmbito internacional, em contextos de temas de interesse comum.

QUADRO 1 - Modelo desconcentrado – principais características.

ARQUITETURA DE INFORMAÇÃO	ARQUITETURA DE TECNOLOGIA DE INFORMAÇÃO
<p><i>Modelo de Captura – desconcentrado, cooperativo, com limitações impostas pelo padrão.</i></p> <p><i>Modelo de Informações/Metadados – definição de documento adequado, padronizado para cada entidade participante, com possibilidade de uso do padrão XML.</i></p> <p><i>Modelo de Comunicação – de largo alcance cooperativo via portal local e Portal integrador; modelo de navegação com Limitações.</i></p> <p><i>Modelo de Operacionalização – de baixo custo relativo, com tecnologias definidas providas pelo gestor.</i></p> <p><i>Gestão de Conteúdos de Informações – desconcentrada.</i></p>	<p><b>Hardware</b> – cada entidade participante provê seu Hardware.</p> <p><b>Software</b> – de baixo custo para o cooperado, com tecnologias definidas e providas pelo gestor. Engenho de Busca com Limitações; usabilidade com restrições.</p> <p><b>Infra-estrutura de Redes</b> – Internet e redes locais.</p> <p><b>Dados</b> – banco de dados proprietário de baixo custo provido pelo gestor.</p> <p><b>Gestão de TI/Segurança</b> – local/desconcentrado/ melhor custo-benefício.</p>

**Modelos concentrados**

O sistema de pesquisa da legislação federal brasileira remete a consultas em três principais fontes no Modelo Concentrado: o Sistema de Informações do Senado Federal – SICON (SICON, 2006), mantido pela Subsecretaria de Informações do Senado (PRODASEN); a Base de Legislação Federal da Presidência da República (Presidência da República, 2006); e a Imprensa Nacional (Imprensa Nacional, 2006).

O Poder Judiciário também tem prosperado em iniciativas importantes no aspecto do gerenciamento eletrônico de processos e jurisprudência. Exemplo disso é o portal do Supremo Tribunal Federal (STF, 2006), sendo o modelo mais adotado pelos órgãos desse poder, havendo inúmeros de utilização de arquiteturas de Modelo Concentrado, com tecnologias similares e proprietárias, em especial quanto aos bancos de dados textuais.

O Tribunal de Contas da União – TCU (TCU, 2007) vem modernizando seus sistemas de acesso em grandes acervos eletrônicos baseados no Modelo Concentrado. Até o final da década de 90, o TCU adotava solução tecnológica de banco de dados textual em ambiente computacional de grande porte (Mainframe), oferecendo ao público interno e ao Congresso Nacional pesquisa textual no inteiro teor de seus Acórdãos e Decisões.

Entre 1998 e 2002 disponibilizou para o público interno (Intranet) funcionalidades para pesquisa em banco de dados textual proprietário com inúmeros recursos, deixando, porém, o público externo (*Internet*) atendido por ferramenta limitada, o que exigia conhecimento das estruturas das bases de dados e das sintaxes de comandos do banco de dados.

A partir do ano de 2003, o TCU contratou no mercado o desenvolvimento de aplicação, com o uso de tecnologia de padrão aberto Java, denominada Portal de Pesquisa Textual. O Portal acessa o banco de dados proprietário em base de dados montada e mantida pelo próprio TCU, sem utilização de bases externas ou convênios com outros órgãos ou empresas. A solução possibilitou – sem a necessidade de treinamento ou

conhecimento das estruturas das bases e dos comandos e sintaxe próprios do banco de dados – a utilização de recursos de pesquisa, tais como refinamento, ordenação, coletânea de documentos, escolha e inserção de operadores textuais e campos para pesquisa, destaque e navegação de termos pesquisados, resumo detalhado de documentos recuperados, *links* entre documentos, apresentação de documentos com formatação rica, dentre outros.

Pode-se citar, também, uma experiência internacional que utiliza o Modelo Concentrado, a Rede Global de Informações Legais (GLIN, 2006), uma iniciativa do Congresso Americano com a rede denominada *Global Legal Information Network* (GLIN), um banco de dados de leis, regulamentos e outras fontes legais complementares:

Os recursos de Tesouro – índice controlado utilizado tanto na indexação, quanto na recuperação de documentos – e o de Vocabulário Controlado são importantes para a qualidade das indexações bem como para sua posterior recuperação.

Como se verifica, o Modelo Concentrado ocorre em organizações com modelos institucionais em geral concentrados, herança esta de padrões tecnológicos, legados da época dos *mainframes*. O padrão de TIC centralizado determina o modelo de informações.

O Modelo Concentrado apresenta vantagens significativas no aspecto de maior governabilidade por parte do gestor. Entretanto, os custos podem ser muito altos, sendo necessário prover todos os meios necessários para sua operacionalização.

No Modelo Concentrado, devido à predominância do uso de padrões proprietários, pode ocorrer uma redução no tempo de vida das aplicações, trazendo uma dificuldade no momento de se garantir a continuidade dos serviços envolvidos.

A estruturação e administração de tratamento e digitalização de informações são mais complexas do que a questão das tecnologias eleitas. No caso específico da legislação e dos atos normativos, dos órgãos e entidades públicas, o tema legislação não é, em geral, o foco principal dessas organizações. Os serviços de tratamento e carga da informação, quando contratados a terceiros, podem ter como consequência a transferência do conhecimento sobre a análise desses atos e da base de conhecimento resultante. As vantagens da flexibilização administrativa, advindas da terceirização, podem se tornar um entrave por ocasião de eventuais rupturas ou da necessidade de renovações contratuais dos serviços de tratamento de informações e da base de conhecimento.

Além disso, é comum empresas com *expertise* em legislação disponibilizarem seus acervos para consulta mediante preço em geral atrativo, na forma de serviços de assinatura a pacotes de mercado, mas, na eventualidade da ruptura contratual, pode haver solução de continuidade e necessidade de acordos sobre a transferência do acervo.

A usabilidade, o fácil acesso por meio de navegadores comuns (*browsers*) e a visualização de textos na íntegra podem se constituir num diferencial em serviços de portal com o uso do Modelo Concentrado em sistemas de pesquisa em acervos eletrônicos.

Os pontos críticos no Modelo Concentrado são: o uso de tecnologias proprietárias e contratos de prestação de serviços completos (acervo, tratamento de informação, *hardware* e *software*, serviços de portal, *call center*, tira-dúvidas, etc.); o processo de gestão, calcado em níveis de serviços contratados e administrados; e as métricas para auferir o retorno sobre os investimentos e o custo-benefício.

As Tecnologias proprietárias, que atendem a inúmeros clientes, podem ser altamente especializadas, mas tendem a um baixo nível de integração institucional, fazendo com que essas aplicações fiquem isoladas do contexto das demais aplicações.

Com tudo isso, devido ao alto grau de especialização, é sem dúvida o modelo mais utilizado por órgãos e empresas com necessidade de soluções rápidas e de custo-benefício razoável – até mais barato do que desenvolver internamente –, na forma de assinaturas a bases de legislação de empresas especializadas no ramo.

Modelos de gestão de conteúdos de forma centralizada têm se popularizado, devido, principalmente, a questões econômicas do negócio. A acirrada competição internacional via globalização, nos últimos cinco anos, acarretou a acentuada queda dos custos das telecomunicações e do processamento centralizado. O modelo desconcentrado pode se tornar antieconômico, comparativamente ao modelo concentrado, quando somados os custos de todos os cooperados, os quais, muitas vezes, não ficam tão evidentes. Percebe-se que as grandes organizações estão preferindo concentrar sua governança em TIC, e, portanto, também a gestão dos recursos de TIC em suas matrizes, de modo a reduzir problemas de comando e controle e, também, os custos de manutenção de equipes especializadas.

## QUADRO 2 - Modelo concentrado – principais características.

ARQUITETURA DE INFORMAÇÃO	ARQUITETURA DE TECNOLOGIA DE INFORMAÇÃO
<p><i>Modelo de Captura – concentrado, não cooperativo, total governabilidade do gestor.</i></p> <p><i>Modelo de Informações/Metadados – definição de documento flexível, padronizado pelo gestor; não requer maiores preocupações com portabilidade ou troca de informações com terceiros.</i></p> <p><i>Modelo de Comunicação – concentrado com maior governabilidade do gestor.</i></p> <p><i>Modelo de Operacionalização – de médio custo, com tecnologias, em geral, proprietárias.</i></p> <p><i>Modelo de Gestão de Conteúdos – concentrada.</i></p>	<p><b>Hardware</b> – totalmente provido pelo gestor.</p> <p><b>Software</b> – em geral tecnologias proprietárias e altamente especializadas contratadas, de médio custo; engenhos de busca especializados; qualidade na aplicação e na navegação.</p> <p><b>Infra-estrutura de Redes</b> – capaz de assegurar um alto volume de usos concorrentes; alto custo de infra-estrutura.</p> <p><b>Dados</b> – banco de dados em geral proprietário (textual); dificuldades de integração com outras bases institucionais.</p> <p><b>Gestão de TI/Segurança</b> – em geral terceirizada; contratação total de serviços (acervo e sistema); melhor custo-benefício.</p>

Embora se utilizado de tecnologias, em geral, proprietárias, as facilidades de busca e o acesso amplo a toda a base, via navegador (*browser*), por qualquer cidadão com acesso à *Internet*, são ganhos significativos desse modelo em relação a plataformas dependentes de um

conjunto de bases de fontes de informações dispersas física e logicamente. O QUAD. 2, a seguir, ilustra as características do Modelo Concentrado.

## Modelos híbridos

Um exemplo de processo cooperativo com o uso de padrões tecnológicos predominantemente abertos é a Comunidade Virtual do Poder Legislativo - INTERLEGIS (INTERLEGIS, 2006). O programa é composto por quatro áreas de atuação: Tecnologia, Informação, Comunicação e Educação. Foi desenvolvido pelo Congresso Nacional, em parceria com o Banco Interamericano de Desenvolvimento (BID), visando à modernização e à integração do Poder Legislativo em seus diversos níveis. Os meios utilizados são as novas tecnologias de informação (*Internet*, videoconferência e transmissão de dados).

Dentre as funcionalidades no INTERLEGIS, algumas operam no Modelo Concentrado, a exemplo das pesquisas on-line em bancos de dados centralizados. Contudo, há processos desconcentrados, com a disponibilização de infra-estrutura onde os integrantes colocam suas informações em páginas próprias na *Internet*, para distribuição geral ou seletiva de documentos, como discursos ou projetos de lei, para o intercâmbio de informações entre as esferas do Legislativo e para acesso às páginas particulares de cada integrante da rede. As informações ficam, assim, disponíveis para qualquer pessoa que tiver acesso à *Internet*.

Tanto o Modelo Concentrado quanto o Desconcentrado apresentam-se como alternativas interessantes, mas podem não ser os mais adequados para o conjunto de acervos expressivos, visto o montante crescente e geograficamente disperso de organizações parceiras. Nessas situações, não basta um portal indicador de onde as bases se localizam, mas sim, tornam-se necessários engenhos de busca e aplicações inteligentes que pesquisem além das estruturas de índices ou de diretórios não estruturados, isto é, nos conteúdos.

Um modelo que consiga integrar dados em bases e suportem um grande volume de dados como também uma expressiva concorrência de usuários (*Very Large Databases - VLDBs*), por meio da *WEB*, mostra-se mais adequado, quando as soluções de arquiteturas tradicionalmente utilizadas não suportam tais volumes.

O problema ocorre, principalmente, devido a grandes massas de dados *on-line*, onde o fluxo de dados entrando, ou de requisições de atualização, pode ser significativo. Tais questões envolvem não só infra-estrutura física, mas, principalmente, uma abordagem lógica para serem operacionalizadas<sup>2</sup>.

---

<sup>2</sup> CAPUANO, Ethel Airton. Arquitetura Tecnológica e Seleção de Sistemas Gerenciadores de Bases de Dados na Previdência Social, relatório técnico interno de nov-2002, onde o autor reporta ao artigo de *Jack G. Conrad et al*, denominado "Database Selection Using Actual Physical and Acquired Logical Resources in a Massive Domain-specific Operational Environment". Hong Kong: 28th VLDB Conference, realizada em 2002.

Exemplo disso é a solução proposta para os serviços *WEB* prestados pelo sistema *Westlaw*<sup>3</sup> (WESTLAW, 2006), com mais de seis *terabytes* de dados textuais, quinze mil bases de dados on-line e 250 mil usuários/dia. Os autores apontam para a criação de uma infra-estrutura que suporte uma seleção automática de bases de dados no serviço de pesquisa mais ampla. Dividem os aspectos de pesquisa distribuída em atividades principais<sup>4</sup>: ordenamento, seleção e busca de coleções de dados e consolidação dos resultados em um conjunto uniforme.

O caso da conhecida enciclopédia "livre" *Wikipedia* (WIKIPEDIA, 2006) é singular. Possui conteúdo, edição e usos aberto e gratuitos, do tipo almanaque, multilíngüe, de criação e desenvolvimento colaborativo e de administração e operação sem fins lucrativos pela Fundação de *Wikimedia*. Sua base de dados contém<sup>5</sup> cerca de um milhão e meio de artigos em inglês, 478 mil em alemão, 374 mil em francês, 304 mil em polonês, 188 mil em português, entre outros. O projeto possui características essenciais: é, ou tende a ser, uma enciclopédia; é um "wiki"<sup>6</sup>, e dessa forma pode ser editada, com algumas exceções, por qualquer pessoa, ou seja, qualquer usuário pode introduzir textos.

A Biblioteca Digital Jurídica do Superior Tribunal de Justiça – DBJur/STJ (STJ, 2007) é o primeiro repositório institucional não acadêmico com utilização de tecnologia "Dspace" ([www.dspace.org](http://www.dspace.org)), desenvolvida pelo Massachusetts Institute of Technology – MIT/USA em parceria com a Hewlett-Packard – HP, baseada em *softwares* livres e gratuitos. Por capitalizar os benefícios dos modelos concentrado e desconcentrado, integração por meio de portais e padrões abertos, pode ser considerada solução híbrida.

A DBJur/STJ é uma rede de bibliotecas digitais formada pelos órgãos do Poder Judiciário, englobando as esferas federal e estadual, além dos órgãos essenciais e auxiliares da Justiça. Pretende integrar, em um único portal, os mais importantes repositórios de informação digital jurídica do Judiciário, de forma a permitir consultas unificadas nesses acervos e possibilitar respostas instantâneas. Cada membro armazena seus documentos em seu banco de dados, propiciando um trabalho cooperativo e um serviço gratuito para os usuários. A padronização da linguagem de alimentação e de recuperação da informação de legislação, doutrina e jurisprudência, através de um Tesouro único, permite buscas e consultas

---

<sup>3</sup> Os dois principais sistemas de bases de dados jurídicas nos Estados Unidos são LEXIS (*Reed Elsevier*) e WESTLAW (West Group). Ambos os serviços oferecem um conjunto de bases em texto completo, cobrindo todo o espectro das publicações jurídicas. Um exemplo dessas bases é o LEGALTRAC (Gale Group), que contém a versão eletrônica do "Current Law Index". Indexa mais de 900 títulos de periódicos especializados, com revisões e boletins a respeito da legislação mundial e americana e fornece texto completo de artigos selecionados (CEDÓN, 2002).

<sup>4</sup> *Citando trabalhos de* Gravano, Callan, French e outros.

<sup>5</sup> *Dados de out. 2006.*

<sup>6</sup> *Um Wiki é uma coleção de páginas WEB conectadas entre si, sendo que qualquer uma delas pode ser visitada e editada por qualquer pessoa em qualquer momento. Mesmo que suas bases sejam centralizadas, opera de modo cooperativo, podendo, no futuro, possuir bases fisicamente descentralizadas, uma vez que se utiliza de padrões abertos e portáveis.*

simultâneas a diversos acervos digitais – decorrentes do Protocolo OAI (*Open Archives Initiative Protocol for Metadata Harvesting*)<sup>7</sup>.

A opção por padrões abertos é uma experiência importante para aqueles que buscam soluções com o melhor custo benefício e padrões tecnológicos mais estáveis, sem que se perca o horizonte da necessidade do bom desempenho final dos serviços.

No aspecto interno de integração de bases de dados com diversas tecnologias, visando minimizar o esforço de captura das informações, utiliza-se uma camada integradora denominada Serviços *WEB* (*WEB Services*), baseada na maneira como as informações são centralizadas, destacando-se como Modelo Híbrido de arquitetura voltado para serviços *WEB*, sendo um padrão cada vez mais utilizado. Os *WEB Services*, portanto, apresentam-se como uma forma simples e padronizada de utilizar a rede mundial para integrar sistemas heterogêneos de forma flexível, substituindo as tradicionais estratégias tecnológicas denominadas *Enterprise Application Integration* (EAI).

O conjunto de metadados, independentemente das tecnologias, define um conjunto amplo de dados estruturados e não estruturados, e acervos documentais, identificando-se as fontes e usos das informações, suportando padrões abertos e portáteis. A forma como as informações são comunicadas é particular de cada aplicação e definida em sua camada. Entretanto, mecanismos de busca potentes selecionam as bases de dados através de amplos serviços de pesquisa, sendo que uma camada adicional, intermediária, integra a camada das aplicações com as diversas bases de dados e serviços.

Os estudos técnicos permitem concluir sobre a importância da integração dos dados. Enquanto os sistemas transacionais<sup>8</sup> legados e os repositórios de dados gerenciais<sup>9</sup> expressam a linguagem de pesquisa (*query languages*), as aplicações de integração adotam o padrão XML (*eXtended Markup Language*).

Sistemas de gerenciamento de conteúdos empregam aplicações para gerenciar e consultar artefatos, tais como documentos, músicas, imagens e vídeos, independentemente deles terem sido manuseados localmente ou gerados por fontes externas, de serem dados estruturados (bancos de dados relacionais) ou não estruturados (hipertexto, bancos de dados textuais).

Os dados, para serem consolidados, podem requerer um processo de limpeza, transformação e extração antes de serem armazenados.

As vantagens de se utilizar essa abordagem são: a simplicidade, sendo mais fácil de operacionalizar do que as soluções tradicionais de integração; o uso de padrões abertos, ao invés de tecnologias proprietárias – embora estas também possam ser utilizadas de modo complementar –; a flexibilidade, onde as alterações nos componentes são

---

<sup>7</sup> Ver <http://www.openarchives.org/OAI/openarchivesprotocol.html#Introduction>. Acesso em: 13 mar. 2007.

<sup>8</sup> São denominados “online transactional process” – OLTP.

<sup>9</sup> São denominados “online analytical process” – OLAP.

mais simples; o escopo, uma vez que cada sistema pode ser tratado como um componente e encapsulado<sup>10</sup>.

A integração de informação no Modelo Híbrido, em especial baseado em *Web Services*, consoante FIG. 1, é uma mescla dos dois modelos anteriores, uma vez que comporta as duas abordagens. A captura dos conteúdos informacionais abrange processos desconcentrados e heterogêneos, intermediados por uma camada que interage com as aplicações, serviços e dados, de forma transparente para o usuário, diretamente ou através do encapsulamento de aplicativos legados.

Assim, na opção de se adotar uma arquitetura tecnológica baseada em *WEB Services* e padrões abertos, tornam-se necessários investimentos no desenvolvimento da camada intermediária (*middleware*) de componentes e operacionalização da arquitetura em camadas.

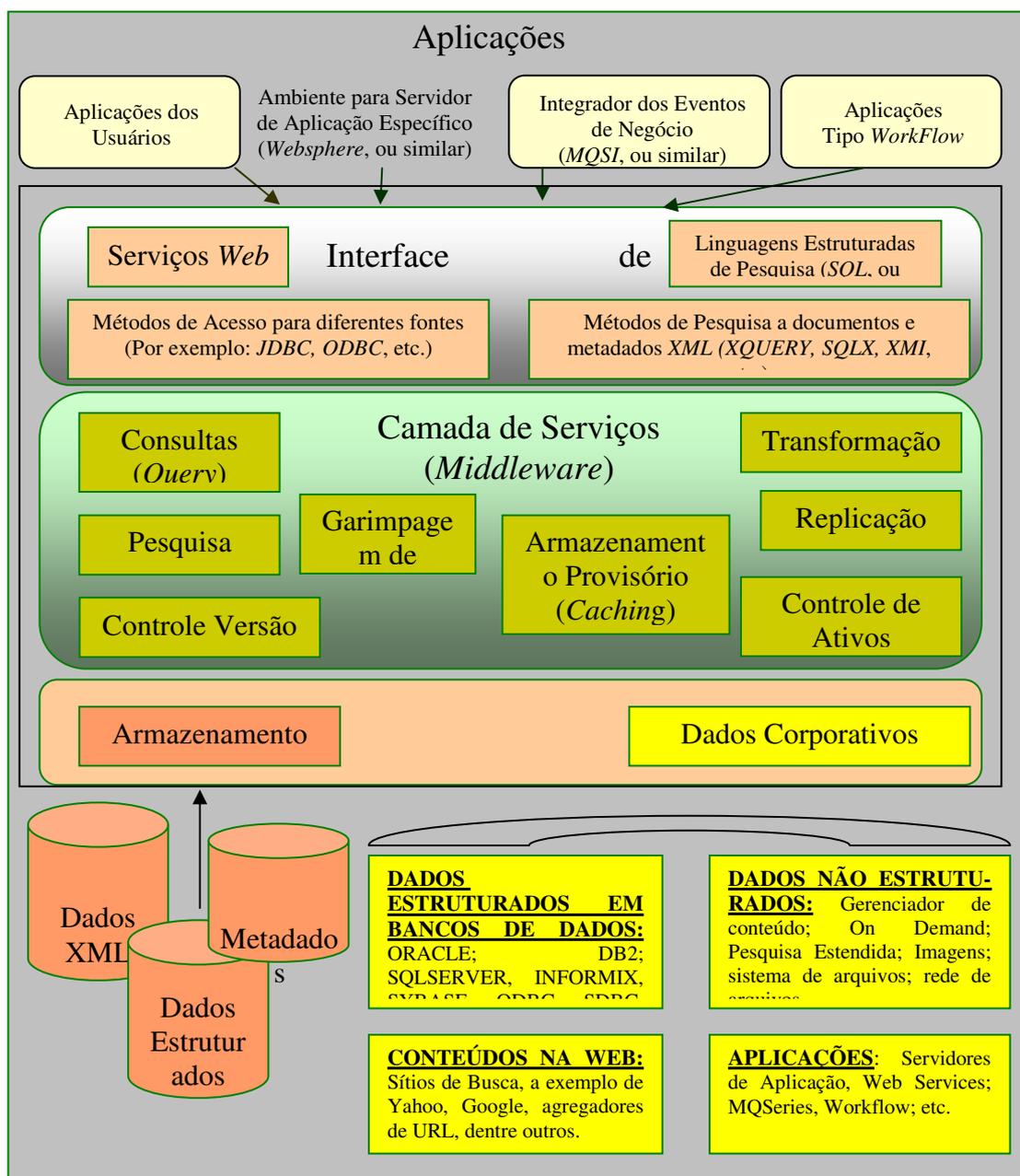
A utilização de sistemas de arquitetura aberta permite o desenvolvimento gradativo e incremental, adequando-se aos orçamentos das organizações e propiciando melhor adaptação de suas funcionalidades, uma vez que os códigos fontes estão disponíveis para alteração e podem ser atualizados no próprio país.

A maioria das atividades de preparação das informações não são significativamente automatizadas, mesmo com a utilização de tecnologias, requerendo pessoal técnico e especializado no tratamento documental.

Um dos maiores desafios na estruturação de *sites* com expressivo volume de textos é a implantação da base inicial, o tratamento e digitalização de documentos, bem como a organização do acervo de normativos, uma vez que a recuperação depende de como as normas se correlacionam, de suas revogações e alterações futuras.

---

<sup>10</sup> Na abordagem tradicional, todos os sistemas são tratados ao mesmo tempo, uma vez que fazem parte da mesma solução monolítica de integração.



O QUADRO 3, a seguir, sintetiza as características do Modelo Híbrido.

<sup>11</sup> Adaptado de WEIGANG, Li; CAPUANO, Ethel Airton; RIECKEN, Rinalda. **Relatório Técnico – Proposições e Projetos para Adequação da Arquitetura Tecnológica da Previdência Social**, derivado de ROTHMary Tork; SCHWARZ, Peter dos autores, From Data Management to Information Integration. A Natural Evolution, de 2002, relatório apresentado à Assessoria Estratégica de Tecnologia e Informação - AETI/SE/MPS, em jun. -2003.

<sup>12</sup> *Observação: as tecnologias citadas são apenas exemplificativas, não significando preferência por marca ou fornecedor.* Adaptado de WEIGANG, Li; CAPUANO, Ethel Airton; RIECKEN, Rinalda. **Relatório Técnico - Proposições e Projetos para Adequação da Arquitetura Tecnológica da Previdência Social**, derivado de ROTHMary Tork; SCHWARZ, Peter dos autores, From Data Management to Information Integration. A Natural Evolution, de 2002, relatório apresentado à Assessoria Estratégica de Tecnologia e Informação - AETI/SE/MPS, em jun. -2003.

□ *Observação: as tecnologias citadas são apenas exemplificativas, não significando preferência por marca ou fornecedor.*

QUADRO 3 - Modelo híbrido – principais características.

ARQUITETURA DE INFORMAÇÃO	ARQUITETURA DE TECNOLOGIA DE INFORMAÇÃO
<p><b>Modelo de Captura – geralmente desconcentrado e heterogêneo.</b></p> <p><b>Modelo de Informações/Metadados – definição de documentos flexíveis, conforme desejado e implementado; repositório de metadados/padrões abertos/XML.</b></p> <p><b>Modelo de Comunicação – flexível, conforme desejado e implementado; interfaces de comunicação a serem desenvolvidas com padrões abertos e engenhos de busca inteligentes.</b></p> <p><b>Modelo de Operacionalização – em camadas, flexível e estável.</b></p> <p><b>Modelo de Gestão de Conteúdos – desconcentrado e heterogêneo.</b></p>	<p><b>Hardware</b> – adequado à estrutura em camadas; bastante distribuído e heterogêneo.</p> <p><b>Software</b> – de custo expressivo inicial, tende a ser mais econômico no longo prazo; integra um mix de tecnologias; engenhos de busca especializados precisam ser desenvolvidos, encapsulados, podendo ser reutilizados; padrões abertos e portáteis;</p> <p><b>Infra-estrutura de Redes</b> – WEB; logística para o middleware (camada intermediária).</p> <p><b>Dados</b> – armazenamento flexível em SGBDs robustos e distribuídos; permite a integração entre dados estruturados e não estruturados.</p> <p><b>Gestão de TI/Segurança</b> – especializada; requer um tempo razoável para o seu desenvolvimento.</p>

Resumo dos três modelos estudados

O resultado do estudo comparativo entre os três modelos de organização de grandes acervos eletrônicos, identificados na pesquisa exploratória, é o demonstrado no QUADRO 4, a seguir, que resume seus principais elementos e características, objetivando, eventualmente, orientar o processo de decisão de gestores na escolha do modelo mais adequado.

QUADRO 4 - Comparativo entre os três modelos

CARACTERÍSTICAS	MODELOS		
	DESCONCENTRADO	CONCENTRADO	HÍBRIDO
Finalidade	.Arranjo colaborativo. Suporte a processos de integração corporativa, entre empresas e entre órgãos governamentais.	.Suporte a processos de trabalho interno e referência para consultas.	.Reúne as características dos outros dois modelos.
Usuários	.Impacto da expressiva quantidade de usuários é minimizado devido à fragmentação das bases.	.Concentração no acesso a um mesmo local por uma grande quantidade de usuários, podendo gerar gargalos.	.Arranjo transparente para os usuários sobre onde as bases estão localizadas fisicamente.
Escopo da Informação	.Amplio, abrangendo também temas correlatos.	.Estreito, restrito aos temas específicos do negócio.	.Amplio, conforme as necessidades.
Estratégia de TIC	.Adota uma proposta colaborativa; e .Favorável no curto e longo prazos.	.Focada nas atividades burocráticas da organização; e .Favorável no curto prazo; desfavorável no longo prazo.	.Adota proposta híbrida; e .Desfavorável no curto prazo, mas favorável no longo prazo.
Tecnologias	.Desenvolvidas pelos patrocinadores / gestores; custo irrisório para os cooperados.	.Proprietárias.	.Plataforma aberta, em camadas.
Esforço para Construção e Operacionalização	.Sistema pouco complexo; e .Esforço distribuído entre os gestores.	.Sistema mais simples; adoção de soluções prontas; e .Não envolve articulação com parceiros.	.Complexo, envolvendo a interação com bases legadas de vários gestores; e .Agenda colaborativa.
Complexidade da Solução	.Baixa complexidade para operacionalizar; complexidade na gestão de bases distribuídas.	.É a solução técnica menos complexa.	.É a solução mais complexa.
Custo de infra-estrutura e suporte	.Custos diluídos entre os gestores.	.Custos concentrados para viabilizar a infra-estrutura de TIC.	.Custos concentrados no início e diluídos no longo prazo.
Complexidade da Gestão	.Gestão complexa.	.Gestão simples	.Gestão local menos complexa.
Qualidade das Pesquisas	.Qualidade Menor.	.Qualidade Maior.	.Qualidade dependente da solução, no longo prazo.

## Conclusão

A acumulação de grandes volumes de dados e informações em meio eletrônico vem tornando necessário o estudo de modelos de organização desses acervos.

Distingue-se a arquitetura de informação, relacionada à modelagem e gestão dos conteúdos e do negócio, da arquitetura de TIC, nos aspectos da gestão e uso das tecnologias de informação das telecomunicações e, portanto, preocupada com a operacionalização da primeira.

A pesquisa exploratória realizada está concentrada em uma classificação de modelos segundo essas duas óticas de arquitetura, uma mais paradigmática e outra mais operacional, a partir de prospecção na *Internet*, destacando-se seus principais elementos e características.

Três modelos empíricos foram então verificados: Desconcentrado, Concentrado e Híbrido.

O Modelo Desconcentrado, em uma avaliação simplificada, é o que apresenta melhor custo-benefício, visto que decorrente de soluções de operacionalização disponíveis no mercado, com a finalidade de organização e cooperação de acervos de informações especializadas.

O Modelo Concentrado é de rápida operacionalização e melhor governança, uma vez que o patrocinador define quais conteúdos deverão ser congregados e quais as tecnologias a serem utilizadas e, de um modo geral, responsabiliza-se pelos investimentos e toda a operacionalização.

O Modelo Híbrido apresenta-se, no longo prazo, como o de características mais flexíveis, com foco na integração e menos dependente de tecnologias para o seu desenvolvimento, embora, no início, requeira maiores investimentos devido à necessidade de se desenvolver a camada integradora.

## Referências

ANVISA. Agência Nacional de Vigilância Sanitária. Disponível em: <http://www.anvisa.gov.br/legis/index.htm> . Acesso: em 12 out. 2006.

BIREME. Centro Latino-Americano e do Caribe de Informação em Saúde. Centro Especializado Organização Pan-Americana de Saúde – OPAS. Disponível em: <http://www.bireme.br/php/index.php>. Acesso em: 12 out. 2006.

BVS/MS. Biblioteca Virtual em Saúde - Ministério da Saúde. Disponível em: <http://bvsmms.saude.gov.br/html/pt/home.html>, ou através do Portal [www.saude.gov.br](http://www.saude.gov.br), selecionando o assunto “Legislação”. Acesso em 12 out. 2006.

GLIN – Global Legal Information Network. Rede Global de Informações Legais do Congresso Americano. Disponível em: <http://www.glin.gov/> . Acesso em: 13 out. 2006.

Imprensa Nacional (Brasil). Sítio oficial da Imprensa Nacional. Disponível em: <http://www.in.gov.br/imprensa/in>. Acesso em: 13 out. 2006.

INTERLEGIS. Comunidade Virtual do Poder Legislativo. Disponível em: <http://www.interlegis.gov.br>. Acesso em: 13 out. 2006.

LIMA-MARQUES, Mamede. Arquitetura de um sistema de informação. Trabalho publicado no Departamento de Informática da Universidade Federal de Uberlândia, 2000.

LILACS. Literatura Latino-Americana e do Caribe em Ciências da Saúde. Disponível em: <http://bvsmodelo.bvsalud.org/php/level.php?lang=pt&component=27&item=3>. Acesso em: 12 out. 2006.

Presidência da República (Brasil). Portal da Presidência da República para consultas a legislação. Disponível em: <http://www.presidencia.gov.br/legislacao/>. Acesso em: 13 out. 2006.

SICON. Portal do Sistema de Informações do Senado Federal – SICON. Senado Federal (Brasil). Disponível em: <http://www6.senado.gov.br/sicon/PreparaPesquisa.action>. Acesso em: 13 out. 2006.

SciELO - Scientific Electronic Library Online. Disponível em: <http://www.scielo.br/scielo.php?lng=pt> . Acesso em 12 out. 2006.

Superior Tribunal de Justiça (Brasil). Sítio para consulta à biblioteca digital jurídica - DBJur. Disponível em: <http://bdjur.stj.gov.br/dspace/>. Acesso em 13 mar. 2007.

Supremo Tribunal Federal (Brasil). Sítio para consulta à base de legislação do órgão. Disponível em: <http://www.stf.gov.br/institucional/resolucoes/ApresentacaoRES.asp>. Acesso em 12 out. 2006.

Tribunal de Contas da União (TCU). Sítio para consulta à base de normas e jurisprudência do órgão. Disponível em: [http://www2.tcu.gov.br/portal/page?\\_pageid=33,210362&\\_dad=portal&\\_schema=PORTAL](http://www2.tcu.gov.br/portal/page?_pageid=33,210362&_dad=portal&_schema=PORTAL). Acesso em 05 fev. 2007.

WWWIsis. Tecnologia utilizada pela BIREME. Disponível em: <http://productos.bvsalud.org/product.php?id=wwwisis&lang=pt>. Acesso em: 12 out. 2006.

Westlaw. Base proprietária de pesquisa a jurisprudência da Thomson West. Disponível em <http://www.westlaw.com>. Acesso em: 13 out. 2006.

Wikipedia. Enciclopédia Livre, em diversos idiomas. Disponível em: <http://www.wikipedia.org/>. Acesso em: 13 out. 2006.