

# A necessidade de integração de fontes heterogêneas de dados em projetos de Gestão do Conhecimento

Mauricio Barcellos Almeida<sup>1</sup>

*Projetos de implantação de gestão do conhecimento<sup>2</sup> em organizações que têm obtido sucesso procuram, não apenas gerenciar os processos através de ferramentas da tecnologia, mas atuar na dimensão de conhecimento neles implícita. A tecnologia da informação é importante para a disseminação de informações, mas deve se tornar parte de um projeto mais abrangente, cujo objetivo é avaliar e tornar disponível o conhecimento organizacional. Apresenta-se a idéia de que uma etapa importante desse projeto trata da solução de problemas relativos à integração das diversas fontes heterogêneas de dados em formato eletrônico presentes em uma organização, para que eles estejam disponíveis de forma transparente para os usuários e possam ser reutilizados. Tomam-se as intranets corporativas como ambiente representativo de heterogeneidade e apresentam-se diversos problemas inerentes à integração em processos organizacionais. Para que se possa perceber a complexidade do assunto, apresentam-se também algumas possíveis abordagens em estudo para a interoperabilidade entre os diferentes tipos de dados.*

**Palavras-chave:** *Gestão do conhecimento, Tecnologia da informação - Interoperabilidade, Fontes heterogêneas de dados - Integração.*

Recebido em: 08/02/2002 - Aceito para publicação em: 27/02/2002

## Introdução

O conhecimento tem sido reconhecido como um novo ativo nas organizações e o único capaz de promover vantagens competitivas sustentáveis em um cenário de competição cada vez mais acentuada. A tecnologia da informação tem proporcionado maior eficiência na disseminação da informação, principalmente a partir dos grandes avanços da área de telecomunicações. Uma abordagem conjunta, que considere, simultaneamente, gestão do conhecimento e tecnologia da informação, é fundamental para que uma organização possa avaliar de forma adequada seu capital intelectual e para que possa torná-lo disponível em todos os seus setores.

Apresenta-se, neste artigo, a proposta de que uma etapa importante para tal abordagem conjunta deva tratar o problema relativo à integração de fontes heterogêneas de dados presentes em uma organização. Na seção 2, é mostrada uma relação entre gestão do conhecimento e tecnologia da informação, discutindo-se os objetivos da implantação de projetos de gestão do conhecimento nas empresas e como as novas tecnologias participam nessa empreitada. Na seção 3, discutem-se

<sup>1</sup>Mestrando em Ciência da Informação pela UFMG, especialista em administração e especialista em engenharia de *software*.

<sup>2</sup>Utiliza-se o termo Gestão do Conhecimento em concordância com a bibliografia consultada e por ser termo em uso corrente, mesmo entendendo que análise mais aprofundada pode questionar sua validade.



problemas relativos à integração de fontes heterogêneas de dados. Apresentam-se as *intranets* como uma ferramenta tecnológica presente nas organizações, representativa da heterogeneidade de dados e discute-se a necessidade de integrá-los. Na seção 4, são identificadas algumas diferentes abordagens para o problema de interoperabilidade, de forma a possibilitar a percepção da complexidade do assunto. A seção 5, trata da necessidade de se realizar trabalhos complementares para se obter uma amostra mais representativa das diferentes abordagens em estudo. Na seção 6 encontram-se as conclusões.

## Tecnologia da informação e gestão do conhecimento

No cenário turbulento que caracteriza o ambiente atual de negócios, questiona-se sobre quais seriam os fatores essenciais para a sobrevivência das organizações. A cada dia, descobre-se que o conhecimento é, hoje, mais do que nunca, decisivo para o sucesso corporativo e que o capital industrial vem sendo substituído por um novo tipo de ativo, o capital intelectual<sup>3</sup>. Os produtos e serviços oferecidos pelas empresas contêm cada vez mais *porções* desse novo ativo, contabilizando grande quantidade de conhecimento. Essa nova realidade tem marcado o período de transição de uma era baseada em ativos industriais para a nova Era da Informação. Conforme sugere STEWART (1998, p. 9):

*“A história é a reinvenção dos negócios, da vida econômica e da sociedade na Revolução da Informação. Da mesma forma como a Revolução Industrial causou grandes estragos no campo e na cidade incipiente, esse novo evento transfigurará e desfigurará tudo o que tocar – e tocará em tudo”.*

As empresas e instituições que mais caminharam no sentido de entender essa nova realidade foram também as primeiras a enxergar que o capital intelectual não é, ainda, adequadamente estudado e organizado. Com a crescente consciência de que apenas a correta utilização do capital intelectual poderá proporcionar à empresa uma vantagem competitiva sustentável, torna-se importante o estudo de meios para a gestão dessa nova forma de ativo.

## Objetivos da gestão do conhecimento nas empresas

Os projetos de gestão do conhecimento têm por objetivo principal proporcionar às empresas vantagens competitivas a partir de uma abrangente exploração dos dados e das informações presentes no ambiente organizacional, mas nem sempre disponíveis para uso. A utilização dessas informações passou a ser vista como estratégica quando as empresas perceberam que eficiência e tecnologia podiam ser igualadas pela concorrência. A fórmula para a sobrevivência, a longo prazo, seria, então, aprender sempre e se preocupar com o uso inteligente do conhecimento. Estas idéias, já presentes em comunidades acadêmicas e empresariais, se popularizaram a partir dos anos 90 com trabalhos como *A Quinta Disciplina*, de Peter Senge.

O termo gestão do conhecimento passou a ser utilizado para representar a nova forma como as empresas procuravam tratar a informação que produziam,

<sup>3</sup> Para XAVIER (1998, p.9), “capital intelectual é o conjunto de conhecimentos possuídos por uma empresa ou instituição e colocado ativamente a serviço de objetivos econômicos”.

conferindo-lhe agora o *status* de ativo. A gestão do conhecimento, lida, assim, com os aspectos críticos para adaptação e sobrevivência da empresa diante de um ambiente de mudança crescente e descontínua. Conforme citam DAVENPORT e PRUSAK (1998, p. 14) “*O conhecimento não é algo novo. Novo é reconhecer o conhecimento como um ativo corporativo e entender a necessidade de geri-lo e cercá-lo do mesmo cuidado dedicado à obtenção de valor de outros ativos mais tangíveis*”.

Segundo FENSEL (2001) a gestão do conhecimento diz respeito à aquisição, manutenção e acesso ao conhecimento de uma organização e tem por objetivo explorar os ativos intelectuais com o objetivo de se obter maior produtividade e maior competitividade.

A *American Society For Training and Development - ASTD* -, citada em XAVIER (1998, p. 99) também define a gestão do conhecimento, relacionando-a com mecanismos para um melhor desempenho das empresas:

*“Gerência do conhecimento - o processo de criação, captação e uso do conhecimento que melhora o desempenho organizacional. A gerência do conhecimento é mais freqüentemente associada a dois tipos de atividades. Um é a documentação e recuperação do conhecimento individual e sua disseminação por intermédio de mecanismos como os bancos de dados corporativos. Inclui também atividades que facilitam as trocas humanas usando-se instrumentos como softwares de partilha de informações, e-mail e Internet.”*

### **O uso da tecnologia da informação em projetos de gestão do conhecimento**

As novas tecnologias proporcionaram mudanças profundas na vida das pessoas e das empresas. Os recursos tecnológicos, atualmente disponíveis, em muito contribuíram para que as empresas atingissem índices de produtividade inimagináveis há pouco tempo atrás. Recursos da informática, aliados aos avanços nos meios de telecomunicações, possibilitaram interação rápida, simples e segura entre organizações e pessoas. Ao possibilitar que o conhecimento fosse disseminado de maneira simples e barata, a tecnologia proporcionou uma importante contribuição à nova necessidade de se administrar ativos intelectuais. Agora, o conhecimento viaja em redes eletrônicas à velocidade da luz. Conforme citam DAVENPORT e PRUSAK (1998, p. 199): “*A maioria das empresas dá seu primeiro passo rumo à gestão do conhecimento na área de tecnologia*”.

Os projetos de gestão do conhecimento que têm obtido sucesso procuram, não apenas gerenciar os processos através de ferramentas da tecnologia, mas atuar na dimensão de conhecimento neles implícita. As novas ferramentas auxiliam em muitas fases dos processos de gestão do conhecimento. Conforme observam DAVENPORT e PRUSAK (1998, p. xi):

*“Depois de cometer dispendiosos erros por ignorar a importância do conhecimento, muitas empresas estão agora lutando para entender melhor o que sabem, o que precisam saber e o que fazer a esse respeito. Algumas dessas organizações equivocadamente presumiram que a tecnologia poderia substituir a qualificação e o julgamento de um trabalhador experiente. [...] A pressuposição de que a tecnologia pode substituir o conhecimento humano ou criar um equivalente para ele tem se revelado falsa. Os progressos da tecnologia, por outro lado, estão entre os fatores que alimentam o interesse no conhecimento e sua gestão”.*

A gestão do conhecimento converge para um ponto no qual são confrontados a cultura administrativa da empresa, seus recursos humanos e os sistemas de informação que ela utiliza. É no conhecimento coletivo que se baseiam as competências essenciais e esse é aprimorado criando-se redes informais de pessoas que realizam trabalhos afins. O objetivo é integrar grupos de profissionais expostos a classes de problemas e tentativas comuns de solução que, através da troca de experiências e informações, aumentem ou refinem o conhecimento coletivo.

O desafio da tecnologia da informação é permitir a identificação, o desenvolvimento e a implementação de sistemas de informação que possam apoiar a comunicação empresarial e a troca de experiências, de maneira a facilitar e incentivar a união das pessoas em redes informais. A tecnologia da informação deve possibilitar que sejam formadas comunidades de trabalho, e não apenas servir para que as pessoas se comuniquem burocraticamente. Nesses aspectos, a tecnologia tem contribuições a fazer, conforme destacam DAVENPORT e PRUSAK (1998, p.172): “... se o apetite, as habilidades e a atenção para com o conhecimento já estiverem presentes na organização, a tecnologia pode expandir o acesso e tornar mais fácil o problema de fazer chegar o conhecimento certo até a pessoa certa no momento certo.”

Dessa forma, pode-se afirmar que uma das grandes contribuições da tecnologia da informação em projetos de gestão do conhecimento está relacionada à disseminação de dados e informações no ambiente organizacional, conforme destacam novamente os mesmos autores (1998, p. 151):

*“ A mais valiosa função da tecnologia na gestão do conhecimento é estender o alcance e aumentar a velocidade da transferência do conhecimento. A tecnologia da informação possibilita que o conhecimento de uma pessoa ou de um grupo seja extraído, estruturado e utilizado por outros membros da organização e por seus parceiros de negócio no mundo todo”.*

### **Integração de fontes heterogêneas de dados disponíveis nas empresas**

Para que seja possível utilizar a tecnologia da informação como parte de um projeto mais abrangente de gestão do conhecimento, cujo objetivo seja avaliar e administrar o conhecimento organizacional, torna-se fundamental uma abordagem que contemple como será feita a integração das diversas fontes de dados heterogêneas disponíveis em formato eletrônico. Assim, será possível ao usuário acessar, de forma transparente, uma variada gama de possibilidades em seu processo de tomada de decisão.

Para BALDONADO e COUSINS (2002), a informação deve ser considerada em termos do contexto maior no qual ela esteja inserida. Esse contexto é o ambiente informacional de redes no âmbito das organizações, onde o problema de integrar fontes de dados heterogêneas torna-se relevante. Os problemas advindos da heterogeneidade estariam relacionados aos usuários, no que diz respeito a tipo de atividades, experiência, estilo, localização geográfica e ferramentas disponíveis; aos repositórios de informação, que diferem drasticamente uns dos outros em gênero, em modalidade, em assunto e em esquemas empregados para descrevê-los (esquemas de catalogação); e, a questões advindas da diversidade de serviços de informação independentes e de sua distribuição.

O universo de tipos de dados, de ferramentas e de tecnologias disponíveis em uma organização torna o problema de prover acesso às fontes heterogêneas de dados, complexo e de difícil solução.

Segundo PAEPCKE (1996), para que um grupo de indivíduos possa tirar proveito do conhecimento disponível em um ambiente organizacional, várias tarefas seriam necessárias, além da tradicional busca por informações. Essas tarefas estariam relacionadas à descoberta e à seleção de recursos, à recuperação dos dados, à interpretação, ao gerenciamento local da informação e ao compartilhamento dos recursos obtidos.

### Fontes de dados em um ambiente organizacional – as intranets

As principais tecnologias em uso em projetos de gestão do conhecimento são a videoconferência, o *groupware*, grupos de discussão, bases de dados *on-line*, CD-ROMs, Internet e *intranets*, sistemas especialistas, agentes inteligentes, *data-warehouse*, *data-mining* e gerenciamento eletrônico de documentos, procurando-se citar as mais importantes. A amplitude da questão que envolve a gestão do conhecimento e o grande número de tecnologias disponíveis tornam-se importantes se se estudar as formas de integração das diversas fontes de dados eletrônicas disponíveis em uma organização.

Um ambiente bastante citado atualmente, quando se trata de fontes heterogêneas de dados, é a Internet, em função da grande variedade, neste ambiente, de tipos de dados, ferramentas, plataformas etc. Ao se aplicar a tecnologia da Internet limitada ao ambiente organizacional, estar-se-á trabalhando em uma *intranet* corporativa.

Para XAVIER (1998), a implantação de uma *intranet*, incluindo possibilidade de acesso por todos os cargos-chave da organização e treinamento para uso adequado, constitui uma estratégia fundamental para a organização que deseja se iniciar na gerência de seu conhecimento.

Segundo FENSEL (2001), um número crescente de empresas tem percebido que as suas *intranets* são valiosos repositórios de informações organizacionais. Pesquisa recente realizada por empresa de consultoria nos Estados Unidos indica uma tendência no sentido do uso de *intranets* pelas empresas. A amostra constituiu de 400 companhias e procura mostrar a utilização de *intranets* corporativas, além de possíveis vantagens desse uso em projetos de gestão do conhecimento. Alguns resultados dessa pesquisa, realizada pela *Collaborative Systems*, apontam para o crescimento na utilização de *intranets* e sua importância para as empresas:

*“Intranets estão crescendo a uma taxa muito rápida porque os usuários perceberam que elas são mais baratas para implementar e mais fáceis de usar do que a maioria dos softwares proprietários. Elas estão sendo usadas para sustentar diversas funções de colaboração e os entrevistados disseram que desejam expandir a utilização de suas Intranets. Que tipo de trabalhos estão sendo executados em Intranets e Internet? O trabalho em grupo é mais freqüente dentro das empresas do que entre empresas. Encontrou-se a maioria das funções de trabalho em grupo sendo suportadas por Intranets. As funções básicas recaem em cinco categorias principais: gestão de documentos, agenda do grupo e horários, gestão de projetos, comunicação, gestão do conhecimento, custos de aplicações e infra-estrutura [...] Os participantes do estudo relataram sua percepção de que o custo para construir uma infra-estrutura na Web é dramaticamente menor que construir a infra-estrutura para um produto proprietário, como o Lotus Notes”.*

As *intranets* são ferramentas bastante representativas da variedade de documentos em um ambiente organizacional e possibilitam a disseminação e o acesso a dados dispersos e heterogêneos. Segundo GAO E STERLING (2000) a gestão do conhecimento na Internet e em *intranets* (as quais utilizam a mesma tecnologia) diz respeito ao uso destes ambientes para publicação de informações, busca, compartilhamento, reuso, integração, suporte ao trabalho em grupo e tomada de decisão.

### **Necessidade de integrar**

Para PAEPCKE et al.(1997), interoperabilidade é um dos problemas mais críticos para administração de dados nas organizações, desde os anos 90, quando o número de sistemas de computadores, repositórios de informações, aplicações e usuários se multiplicou a uma taxa explosiva.

Os dados nas empresas podem estar em formatos de texto, vídeo ou áudio, com grau variável de estrutura (dados estruturados, por exemplo bancos de dados relacionais; dados semi-estruturados, por exemplo *e-mails* e paginas HTML; dados não estruturados, por exemplo texto livre). Dentre as diversas possibilidades de formatos de dados eletrônicos nas organizações, devem-se destacar os documentos, em forma de arquivos, aqui considerados como entidades representativas de parte do conhecimento organizacional armazenado em computadores. Segundo GAO e STERLING (2000), muitos desses documentos são escritos para leitores humanos, e o conhecimento que eles contém, não pode ser acessado facilmente por programas de computador. Dessa forma, o processo de integração automatizada, na grande maioria dos casos, não é simples.

FENSEL (2001) afirma que dados em estado bruto, em grande quantidade não podem por si só solucionar problemas de negócios ou prover vantagem competitiva. Informação é inútil sem a compreensão de como aplicá-la efetivamente. Dessa forma, pode-se afirmar que é importante a tarefa de integrar e disseminar os dados como forma de lhes conferir utilidade no ambiente corporativo.

O problema da heterogeneidade de fontes de dados é fundamental para o estudo da gestão do conhecimento na empresa visto que hoje, grande parte dos documentos e da sistematização de processos encontra-se em meio digital. Esperar por uma possível padronização, o que não parece provável, deixaria em aberto o problema sobre como integrar os chamados *sistemas legados*. Geralmente desenvolvidos no passado, estes sistemas foram construídos em linguagens e padrões já em desuso, para os quais, muitas vezes, não se encontram, atualmente, profissionais qualificados ou os próprios autores para execução da manutenção.

### **Alguns exemplos de abordagens para os estudos de integração e interoperabilidade**

A tarefa de integrar fontes de dados heterogêneas tem sido bastante estudada atualmente, visto a importância, para as instituições e empresas, do acesso e manutenção do conhecimento que possuem, de forma rápida e simples. A competitividade e a capacidade de reação das empresas no ambiente de negócios



*Interfaces:* KAMIYA, ROSCHEISEN e WINOGRAD (1995) adotam uma abordagem baseada em uma interface que faz a mediação entre vários tipos de recursos de trabalho em grupo como, por exemplo, *e-mail*, índices hierárquicos, recursos compartilhados. Entretanto, o estudo parece ser direcionado para uso apenas na Internet.

*Metadados:* BALDONADO, CHANG e GRAVANO (1998) procuram solucionar problemas de interoperabilidade em ambientes heterogêneos através de uma arquitetura de metadados que integra informações sobre as diversas coleções e serviços disponíveis em uma rede. A arquitetura proposta tem quatro componentes básicos: um modelo de atributos de extratores de dados, um modelo de atributos de tradutores, utilitários para busca de extratores e repositórios de metadados.

*Linguagens de consulta:* GHANG, MOLINA e PAEPKE (1996) abordam problemas de busca em fontes heterogêneas do ponto de vista das dificuldades em integrar diferentes estilos e linguagens de consulta booleanas. Propõem uma arquitetura e um mecanismo para tradução de consultas em uma linguagem comum.

Inúmeros outros estudos e diferentes possibilidades continuam sendo estudadas para a integração de dados e interoperabilidade. Os exemplos e pesquisadores aqui apresentados dão uma mostra da diversidade e complexidade do assunto.

76

## **Trabalhos futuros**

Como trabalho futuro e complementar a este artigo pretende-se elaborar uma pesquisa mais abrangente sobre os principais projetos de integração de fontes heterogêneas de dados e interoperabilidade, de forma a construir um referencial teórico necessário para aplicação em projetos de gestão do conhecimento em empresas.

## **Conclusões**

O problema de integração de fontes heterogêneas em projetos de gestão de conhecimento se apresenta como mais um desafio que confronta a utilização da tecnologia da informação e o conhecimento produzido pela organização. Entretanto, os estudos relativos à integração de fontes heterogêneas têm demonstrado que se trata de um problema complexo, o qual deve ser abordado de diferentes formas. É importante, numa primeira etapa, definir que nível de interoperabilidade é desejado. Assim, poder-se-á buscar um ambiente no qual a informação esteja acessível e seja reutilizada por toda a organização, em diferentes períodos de tempo e independentemente do formato em que esta armazenada. Isso pode possibilitar que um problema da organização resolvido hoje, esteja disponível para outro indivíduo em outra situação, em outra localização geográfica e em outro espaço de tempo. O conhecimento poderá, assim, ser disseminado de forma eficiente e os indivíduos da empresa poderão acessar de maneira simples a parcela de conhecimento armazenada em meio digital e, talvez, tratar de sua gerência.

### **The need for heterogeneous data sources integration in knowledge management projects**

*Knowledge management projects, which have been successfully implemented in organizations, try not only to manage technological tools, but also to act in the knowledge dimension. Information technology is important to the dissemination of which has the objective of making organizational knowledge available. We present the idea that one important stage of this project deals with problems related to the integration of the several heterogeneous data sources found in electronic formats in any company, so as to make them available, clearer and reusable to users. Corporate intranets are taken as examples of representative environments of heterogeneous data sources and several inherent problems related to the organizational processes are also presented. We also present possible approaches to the interoperability between distinct data sources.*

**Key-words:** Knowledge management, Information technology - Interoperability, Heterogeneous data sources - Integration.

## Referências

- BALDONADO, Q.W.M.; COUSINS B. S. *Addressing heterogeneity in the networked information environment*. Disponível na Internet via WWW. URL <http://citeseer.nj.nec.com/27530.html>. Arquivo capturado em 25/01/2002.
- BALDONADO, Q.W.M.; CHANG, C.C.K.; GRAVANO, L. *Metadata for digital libraries: architecture and design rationale*. Disponível na Internet via WWW. URL <http://www.diglib.stanford.edu/cgi-bin/WP/get/SIDL-WP-1997-0055>. Arquivo capturado em 15/02/2002.
- CHANG, CC.K.; GARCIA-MOLINA, H.; PAEPCKE, A. *Predicate rewriting for translating boolean queries in a heterogeneous information systems*. Disponível na Internet via WWW. URL <http://citeseer.nj.nec.com/chang96predicate.html> Arquivo capturado em 15/02/2002.
- COLLABORATIVE STRATEGIES PUBLICATIONS. *Electronic Collaboration on the Internet and Intranets*. Capturado em 26/09/1999. On-line. Disponível na Internet via WWW. URL: [www.collaborate.com](http://www.collaborate.com).
- DAVENPORT, T.; PRUSAK L. *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. Rio de Janeiro: Campus, 1998.
- FENSEL, D. *Ontologies: a silver bullet for knowledge management abd electronic commerce*. New York: Springer-Verlag Berlin Heidelberg, 2001.
- GAO, X.; STERLING, L. Semi-structured data extraction from heterogeneous sources. In: SCHWARTZ, D.G.; DIVITINI, M.; BRASETHVIK, T. *Innovative Internet Information Systems Workshhop - Internet-based organizational memory and knowledge management*. New York: Ideia Group Publishing. 2000.
- KAMIYA, K.; ROSCHEISEN, M.; WINOGRAD, T. *Grassroots: a system providing a uniform framework for communicating, structuring, sharing information and organizing people*. Disponível na Internet via WWW. URL <http://pcd.stanford.edu> . Arquivo capturado em 30/01/2002.
- KESHAV, R.; GAMBLE, R. *Towards a taxonomy of architecture integration strategies*. Disponível na Internet via WWW. URL <http://www.seat.utulsa.edu/> . Arquivo capturado em 30/01/2002.
- PAEPCKE A. Digital libraries: searching is not enough. *D-Lib Magazine*, May 1996. Disponível na Internet via WWW. URL <http://www.dlib.org/dlib/may96/stanford/05paepcke.html>. Arquivo capturado em 20/01/2002.
- PAEPCKE A. et al. *Toward interoperability in digital libraries*. Disponível na Internet via WWW. URL <http://www.diglib.stanford.edu/diglib/pub/reports/>. Arquivo capturado em 26/01/2002.
- PAEPCKE A. et al. *Interoperability for digital libraries: problems and directions*. Disponível na Internet via WWW. URL <http://citeseer.nj.nec.com/119879.html>. Arquivo capturado em 26/01/2002.
- PAEPCKE A. et al. *Search middleware and the simple digital library interoperability protocol* (long version). Disponível na Internet via WWW. URL <http://citeseer.nj.nec.com/paepcke00search.html>. Arquivo capturado em 15/02/2002.
- STEWART, T. *Capital intelectual: a nova vantagem competitiva das empresas*. Rio de Janeiro: Campus, 1998.
- TENANT, R. Interoperability: the holy grail. *Library Journal*, July 1998. Disponível na Internet via WWW URL <http://www.ljdigital.com>. Arquivo capturado em 30/01/2001.
- XAVIER, R. A. P. *Capital intelectual – administração do conhecimento como recurso estratégico para profissionais e empresas*. São Paulo: STS Publicações, 1998.

\*O autor deste artigo agradece as ideias e o apoio do Prof. Dr. Marcello P. Bax, que em muito contribuíram para a melhoria deste trabalho.