

DEPARTAMENTO DE INFORMÁTICA
SETOR DE DOCUMENTAÇÃO
E INFORMAÇÃO

Pontifícia Universidade Católica do Rio de
Janeiro
Rua Marquês de S. Vicente, 225 - casa B
Gavea - RJ - Cep. 22453

DIVISÃO DE INTERCÂMBIO E EDIÇÕES

M 3813

SETOR DE DOCUMENTAÇÃO E INFORMAÇÃO	
CÓDIGO / REGISTRO 4347	DATA 6/7/79
DEPT. DE INFORMÁTICA	

DIRETÓRIO DE DADOS UTILIDADE E
ESPECIFICAÇÃO

Antônio Luz Furtado
Arndt von Staa

Mcc 11/78

INF/001

JAN/79

Série: Monografias em Ciência da Computação

Nº 11/78

Editor da Série: Michael F. Challis

DIRETÓRIO DE DADOS
UTILIDADE E ESPECIFICAÇÃO

por

Antônio Luz Vurtado

Arndt von Staa

Sumário

1	Introdução	1
2	Conteúdo dos Diretórios de Dados	5
3	Criação de um Diretório de Dados	8
4	Usos de Diretório de Dados	11
5	Manutenção do Diretório de Dados	14
6	Requisitos Funcionais de Diretórios de Dados	14
7	Conclusões	16
	Referências Bibliográficas	17
	Apêndice A: "Lay Out" do Diretório de Dados	18

Resumo

Dados são recursos valiosos. O valor dos dados cresce à medida que eles tenham maior qualidade, maior compreensibilidade por parte de seu usuário e maior intercambiabilidade entre os diversos sistemas. Apresentamos aqui o que vem a ser um diretório de dados e analisamos como este contribui para elevar o valor dos dados.

Palavras Chaves:

Dados, Bancos de Dados, Sistemas de Informação, Diretório de Dados; Meta-banco de Dados; Administrador de Dados; Sistemas de Gerência de Bancos de Dados.

Abstract

Data is a valuable resource. The value of data increases as its quality improves, as it becomes more meaningful to its end user, and as it becomes easier to move it between different systems. In this text we discuss what a Data Directory is and how it contributes to the increase in value of data.

Key Words

Data, Data Bases, Information Systems, Data Directory, Meta-Data Base, Data Administrator, Data Base Management Systems.

1 Introdução

Dados são recursos de alto valor, apesar deste ser de difícil quantificação. O valor dos dados depende em grande parte da facilidade com que se tem acesso a eles, os fatores de qualidade (consistência, precisão, completeza) e a facilidade com que são compreendidos e convertidos em informação pelo seu usuário.

Em qualquer empresa ou instituição são manipulados dados, ou seja, existem um ou mais sistemas de informação. Estes dados são arquivados de uma maneira ou outra, seja em arquivos de pastas, em livros ou em arquivos de computador. Em muitas ocasiões o mesmo dado é armazenado em locais diferentes o que poderá trazer diversos problemas, entre eles o de falta de consistência.

Quando dados são armazenados em diversas cópias em arquivos de computador é imprescindível que a representação e o manuseio obedeça a certas normas pois, de outra forma, teremos quase que certamente dados inconsistentes e dificuldades de troca de dados entre os diversos sistemas.

Em muitos casos, organizações já possuem alguns sistemas automatizados, estando no processo de desenvolvimento global envolvendo todos os sistemas. É desejável, portanto uma evolução gradativa tendo por objetivo final o sistema integrado [Canning73]. Esta evolução cautelosa pode ser obtida a partir da normalização da representação e do manuseio dos dados garantindo-se que esta norma seja obedecida por todos os sistemas em existência ou em desenvolvimento.

O diretório de dados é nada mais do que esta normalização do formato e manuseio dos dados [Uhrowczik73, Plagman72, Bontempo73] cabendo ao administrador de dados [Newel73, Fox77] garantir que as normas sejam mantidas, atualizadas e efetivamente adotadas.

Sistemas apoiando diretórios de dados permitem, também, que diversas tarefas rotineiras de codificação possam ser automatizadas. Assim o manuseio dos dados pelos sistemas de programação em todos os sistemas implantados.

Um Sistema Gerência de Bancos de Dados (SGBD) compreende de um conjunto de arquivos. Entretanto representa mais do que simplesmente esse conjunto, como veremos mais adiante.

Em todo o desenvolvimento deste trabalho utilizaremos um exemplo, com a finalidade de tornar mais intuitivos os conceitos a serem discutidos. O exemplo envolve a área de pessoal de uma firma que executa certos projetos de engenharia.

Os tipos elementares de informações, que chamaremos de ítems, incluídos no exemplo, são:

- N - nome do empregado
- S - salário
- C - classe profissional
- H - habilitação
- T - tarefa
- P - projeto
- L - líder de projeto

Esses ítems figuram nos seguintes arquivos:

EMP(N,S,C) - nome, salário e classe profissional de cada empregado

REQ(T,H) - requisito de habilitação para o desempenho de uma tarefa

ATR(N,T,P) - atribuição de tarefa a empregado dentro de projeto

CHF(P,L) - chefia de projeto por um líder

CAP(N,H) - empregado capacitado quanto a uma habilitação

Cada um desses 5 arquivos é um conjunto de registros. Assim, se o registro (P3, Miranda) figura no arquivo CHF, isso significa que o Projeto P3 é liderado por Miranda.

A figura 1 acentua o fato, característico de um SGBD, de que os arquivos possuem ítems em comum.

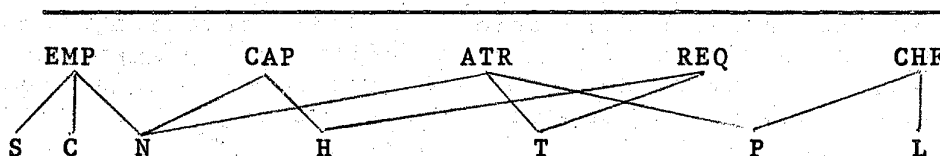


Figura 1. Relações entre arquivos e ítems

Pode-se dizer que o uso de um SGBD não se justifica a menos que se verifique o fato de existirem itens em comum entre arquivos. A existência de dados comuns a vários arquivos permite, por exemplo, que se responda a perguntas como:

"Quais os empregados subordinados a Miranda?"

que não poderia ser respondida consultando um único arquivo. No caso, temos de associar os arquivos CHF e ATR, descobrindo em CHF que Miranda lidera o projeto P3, procuramos então no arquivo ATR todos os registros contendo esse valor P3 e anotamos os nomes dos empregados que aparecem nesses registros.

Outra característica de um SGBD, são certas restrições (1) envolvendo mais de um arquivo. No exemplo, os empregados que trabalham em um projeto (arquivo ATR), devem ter sido previamente admitidos pela firma dentro de uma classe profissional e com um salário (arquivo EMP). Num caso como este torna-se vital a existência de um diretório de dados em particular se estes arquivos não forem mantidos de forma interligada. Isto decorre do fato da atualização em um arquivo poderem requerer que outros arquivos sejam atualizados prévia ou simultaneamente.

Mesmo no caso de arquivos isolados já é um problema vital a autorização de pessoas ao uso dos arquivos. Utilizando-se um SGBD o problema é ainda mais sério porque o uso dos arquivos por diversas pessoas é grandemente estendido. Conseqüentemente, devem ser caracterizados os usuários do sistema, especificando que operações de consulta e/ou atualização são autorizados a executar. No nosso exemplo, os usuários seriam:

- o gerente de pessoal
- o gerente de engenharia
- o gerente de treinamento
- os líderes de projetos

Temos de conciliar o uso do sistema entre esses usuários, de modo que eles possam obter e manipular informações de acordo com suas necessidades, mas sem interferir uns com os outros. Poderíamos definir regras como: só o gerente de pessoal pode admitir um empregado (inserção em EMP) ou demití-lo, mas a demissão só poderá ser efetivada se o empregado não estiver no momento ligado a projeto algum (arquivo ATR), sendo o estabelecimento ou eliminação desta ligação da competência do gerente de engenharia, etc.

(1) Geralmente estudadas como requisitos de integridade ou de coerência na literatura.

Um SGBD também envolve diversas funções administrativas ou auxiliares, tais como restauração de arquivos em caso de erro, contabilização e auditoria do uso dos arquivos, estatísticas sobre o uso, etc. A obtenção de estatísticas sobre o uso é fundamental para a revisão das técnicas de programação, armazenamento e acesso usadas, porque permitem detetar as possíveis ineficiências operacionais do sistema automatizado.

Assim um SGBD é bem mais do que um mero conjunto de arquivos e rotinas de acesso a dados contidos nesses arquivos. Para caracterizar um banco de dados específico precisamos, entre outras, dispor de informações sobre as características que acabamos de considerar. Ou seja: precisamos de um banco de dados sobre o banco de dados. Esse meta-banco de dados é parte do diretório de dados.

Tanto o diretório de dados, quanto o SGBD fazem parte de um conjunto - Banco de Dados Institucional - ao qual pertencem:

- a. O banco de dados - repositório de todos os dados utilizados pela instituição. Isto pode ser implementado através de um conjunto de arquivos independentes e/ou interdependentes (i.é. um banco de dados propriamente dito)
- b. O diretório de dados - o repositório de todos os dados que descrevem o banco de dados, tais como características, responsabilidades, autorizações de acesso, etc.
- c. o administrador de dados - uma função não computarizada responsável pela coordenação de todas as atividades relacionadas com dados, em particular pela aderência às normas estabelecidas para a utilização e o manuseio de dados
- d. O sistema de gerência de bancos de dados - uma função de software que perfaz o armazenamento, recuperação e manutenção dos dados
- e. A interface usuário/sistema - os subsistemas que permitem a usuários diversos orientar o sistema no sentido de que produza os dados desejados.

O aspecto mais importante a ser discutido neste trabalho é que, mesmo que a empresa não disponha de um SGBD no sentido próprio, um diretório de dados pode ser constituído para proporcionar uma visão integrada dos sistemas de informação, cujos dados poderão estar armazenados em diversos arquivos centralizados ou não, computarizados ou não.

2 Conteúdo dos Diretórios de dados.

Nesta seção vamos motivar o conteúdo do diretório de dados. No Apêndice A, é especificado completa e detalhadamente o conteúdo do diretório de dados.

Neste trabalho vamos supor que um diretório de dados é um arquivo de computador. Não entraremos em detalhes sobre as possíveis maneiras de organizar fisicamente esse arquivo. No entanto, é importante notar que ele deve registrar as informações em dois formatos:

- formato legível por programas
- formato legível por pessoas

Quando o gerente de pessoal aciona um programa para demitir um empregado (eliminação de registro do arquivo EMP) este programa seria interceptado por uma rotina do SGBD que verificaria no diretório de dados se esta operação deveria ser precedida de algum teste. A rotina interceptadora encontraria no diretório a indicação de que a operação depende da não ocorrência do nome do empregado em registros de ATR. A rotina interceptadora efetuaria então, uma pesquisa preliminar nesse arquivo. Este é um caso em que os dados contidos no diretório de dados são lidos por programa e portanto supõe-se que estejam armazenados de forma compacta.

Por outro lado, o gerente de pessoal poderia querer saber em que condições poderia demitir um empregado. Para isso consultaria o diretório de dados, através de alguma rotina do SGBD, e obteria em seu terminal uma mensagem do tipo:

"Somente empregados não ligados a projetos podem ser demitidos".

e para que seja possível a produção desta mensagem, supõe-se que ela esteja armazenada no diretório de dados, sob forma textual.

O diretório de dados deve conter descrições relativas ao significado de cada item e dos possíveis interrelacionamentos dos itens, como por exemplo:

- quais os arquivos integrantes do SGBD
- quais são os itens de cada arquivo
- quais são os agregados de itens. Por exemplo, o item (agregado) endereço é formado normalmente pelos itens rua, número, apto. etc.

- qual o significado de cada arquivo. Por exemplo, o arquivo REQ indica habilitações exigidas para o desempenho de tarefas.
- qual o significado de cada ítem em cada arquivo. Por exemplo, o ítem H (habilitação) no arquivo REQ é a habilitação exigida para uma tarefa, enquanto que H no arquivo CAP indica habilitação que um empregado tem.
- qual o significado de cada agregado em cada arquivo.
- qual a necessária compatibilidade entre ítems de arquivos. Por exemplo os líderes L contidos no arquivo CHF são também empregados e, portanto, devem também aparecer como nomes de empregado N no arquivo EMP.
- qual a necessária compatibilidade dependente de conversão (unidade de medida). Por exemplo, no arquivo EMP, os salários (S) podem estar armazenados como salário-mês (para a classe profissional engenheiro) ou como salário-hora (para guindasteiro).
- quais as operações permitidas sobre cada arquivo. Por exemplo, a eliminação de registros pode ser permitida para o arquivo EMP mas não para o arquivo REQ.
- quais as condições e conseqüências de cada operação. Por exemplo, uma demissão de empregado (eliminação de registro no arquivo EMP) pode ser considerada somente se o empregado não estiver ligado a projeto (teste do arquivo ATR), e se executada pode causar o cancelamento de suas habilitações (eliminação de registro no arquivo CAP).
- quais os usuários autorizados a executar cada operação. Por exemplo o gerente de treinamento pode inserir novos registros e alterar registros existentes no arquivo CAP, o líder de um projeto pode consultar o arquivo CAP quanto aos empregados trabalhando em seu projeto, etc.

O diretório de dados também deve conter informações relacionadas com as atividades administrativas a cargo de rotinas do SGBD, como por exemplo:

- condições e procedimentos de restauração de arquivos.
- periodicidade para a confecção de cópias de arquivos para eventual restauração.
- arquivos auxiliares e procedimentos para a prevenção de acidentes.

O diretório de dados deve conter informações a respeito da providência e autorização de acesso dos dados, como por exemplo:

- quem é responsável pela coleta de dados.
- quais os formulários em que aparece o item como dado de entrada e/ou atualização.
- qual o grau de precisão que o dado deve ter. Por exemplo, em um sistema para distribuição de correspondência (mailing list) os endereços tem que estar corretos, enquanto que em um sistema de pessoal alguns erros de endereços podem ser tolerados.
- quem tem o direito de criação sobre um item.
- quem tem direitos de alteração sobre um item.
- quem tem direitos de destruição sobre um item.
- quem tem direitos de acesso sobre um item.
- qual o nível de confidencialidade com que um item deve ser tratado.

Finalmente o diretório de dados deve fornecer indicação sobre o valor dos dados, como por exemplo:

- qual o valor atribuído ao dado (em moeda, ou em outra medida). O valor deve ser discriminado quanto ao dado estar correto, incorreto ou incompleto, ou inexistente.
- qual o custo de obtenção do dado.
- qual o custo de manutenção do dado.
- qual o benefício que alguém não autorizado poderia auferir acessando ou adulterando o dado.

3 Criação de um Diretório de Dados.

Em geral decide-se que informações conterã um diretório de dados, e se obtêm várias dessas informações durante a fase de projeto do banco de dados. Por exemplo, quais serão os arquivos se decide durante a fase de projeto. Por outro lado quais estatísticas de uso a serem anotadas é decidido durante a fase de projeto, porém os dados estatísticos propriamente serão obtidos durante a fase de uso efetivo.

Em muitos casos, como já foi mencionado, é desaconselhável ou mesmo impossível do ponto de vista prático, projetar-se um SGBD diretamente. Nestes casos chega-se a ele por evolução. Necessita-se então um mecanismo normalizador que controle a evolução dos sistemas já existentes de tal forma que os arquivos comecem a tornar-se compatíveis. A criação de um SGBD torna-se então menos complicada e menos traumática para a instituição.

Além de ser difícil a criação de um SGBD, isto é bastante demorado e a instituição nem sempre pode ficar aguardando a criação do SGBD para então partir para a produção dos programas de aplicação. Durante este período, um diretório de dados serve como mecanismo para aumentar a integrabilidade futura dos sistemas de aplicação desenvolvidos em paralelo com a criação do SGBD.

Finalmente, mesmo que não seja desejado criar eventualmente um SGBD, o diretório de dados poderá vir a ser útil no sentido de homogeneizar a nomenclatura utilizada na instituição e, também, para permitir possíveis interações entre arquivos independentes.

Voltaremos agora nossa atenção inteiramente para a obtenção de alguns dados que indenpendem da existência concreta de um SGBD.

Alguns desses dados mais importantes são:

- quais os arquivos existentes na empresa.
- quais os itens de cada arquivo.
- qual o significado de cada arquivo.
- qual o significado de cada item em cada arquivo.

Aos quais acrescentamos, supondo justamente que os arquivos estão dispersos, pertencendo a determinados usuários utilizando-os dentro de sistemas específicos:

- identificação do usuário proprietário, e localização do arquivo. Por exemplo, para o arquivo CAP, o usuário seria, o gerente de treinamento, residente na cidade C.

- identificação do sistema ou sistemas em que o arquivo é usado. Por exemplo, o arquivo CAP seria utilizado nos sistemas de treinamento e de controle de recursos humanos.
- significado desses sistemas. Por exemplo, o sistema de treinamento determina o planejamento do treinamento dos empregados em face das necessidades da empresa para a realização das tarefas dos projetos (1)

O levantamento desses dados costuma ser bastante difícil além de trabalhoso. Vamos supor que ele esteja a cargo de um órgão central de Coordenação dos Sistemas de Informação (CSA). Além disso os vários usuários dentro da empresa, ou possuem suas próprias instalações de processamento de dados, ou utilizam serviços de algum centro de processamento de dados da própria empresa ou externos a esta.

A primeira tarefa é convencer os usuários da utilidade do levantamento, tópico que veremos no item 4 deste trabalho, e obter sua colaboração para o levantamento seja realizado.

A remessa de questionários é desaconselhada porque resultaria em falta de padronização nas respostas, e até na falta parcial ou total de respostas. Deve-se utilizar um formulário, mas encarregar o próprio pessoal responsável pelo levantamento de ir pessoalmente a cada usuário e preencher o formulário por meio de entrevistas. Convém que o formulário tenha uma disposição que facilite sua digitação (perfuração de cartão ou entrada por terminal).

O levantamento produzirá uma informação hierarquizada:

identificação de cada usuário

por usuário: sistemas - identificação e significado

por sistema: arquivos - identificação, significado e local

por arquivo: itens - identificação e significado

A grande dificuldade no levantamento está em como exprimir os vários significados de forma totalmente inteligível, pelo menos para o pessoal da empresa. Os especialistas em programação tendem a atribuir nomes a sistemas, arquivos e itens que não são nada expressivos. Por exemplo, um item poderia ser chamado de X ou mesmo CÓDIGO. No caso do nome do item ser CÓDIGO, este poderia ser código de projeto ou de habilitação, etc.

(1) O sistema dado como exemplo envolveria vários outros arquivos, como por exemplo o arquivo REQ cujo proprietário "natural" seria o gerente de engenharia.

Alguns nomes de itens podem até parecer precisos mas significar coisa totalmente diferente do que parecem. Considere, por exemplo, um item denominado PRODUTO. Pareceria algum material fabricado, negociado ou utilizado pela empresa, no caso porém em um caso observado na prática era o resultado da multiplicação de dois outros itens.

Até nomes realmente claros podem necessitar de explicação. Por exemplo, QUANTIDADE-DE-MATERIAL poderia referir-se a material encomendado, ou existente em estoque, ou consumido, etc.

Mais delicado é o fato de que um mesmo item poder receber identificações diferentes dadas por usuários diferentes. Ao examinar os significados anotados é que o pessoal da CSI irá suspeitar de que se trata do mesmo item, devendo então, voltar a procurar os usuários para verificar essa possibilidade. Se positiva, é então conveniente que a CSI atribua uma identificação padrão ao item e a divulgue pelos usuários. No nosso exemplo, identificações como ATRIBUIÇÃO, CARGO, MISSÃO, etc. poderiam todos denotar o que chamamos de TAREFA.

As inconsistências e deficiências de nomenclatura devem ser eliminadas. Esta tarefa é bastante trabalhosa e requer:

1. a adoção de um estilo padrão para a redação, da definição dos itens. Desta forma facilita-se descobrir itens que possuam o mesmo significado e nomes diferentes.
2. a adoção de nomes padrão, do nome por extenso que deve ser o mais auto-explicativo possível e, do nome de código, a ser utilizado nos programas. Deve ser seguida uma regra de geração do nome de código a partir do nome por extenso. Desta forma os programas tornar-se-ão mais compreensíveis e, conseqüentemente, mais manuteníveis.

O resultado desta tarefa é então algo parecido a um dicionário onde aparecem os verbetes (nome por extenso de itens) e suas definições. Deve-se observar, porém, que os verbetes não possuem sinônimos nem homônimos e, finalmente, as definições devem ser precisas, compreensíveis, completas e não ambíguas.

Somente após esta fase é que convém definir formatos, organizações internas, responsabilidades, etc.

Dificuldades menores no levantamento são, por exemplo, decidir se o levantamento deve abranger somente os arquivos permanentes ou também os temporários, os intermediários, etc., e se deveria ser aproveitada a oportunidade para levantar outros dados como a organização física dos arquivos.

A falta de uniformidade no preenchimento pode causar problemas para as futuras apurações. Por exemplo, se anotarmos a periodicidade de atualização de um arquivo, é preciso ou usar sempre a mesma unidade de tempo ou especificar a unidade usada, como 3 vezes por mês, 2 vezes por ano. Se anotamos o período de retenção temos que em alguns casos o arquivo não é jamais abandonado. Neste caso não devemos algumas vezes anotar que a retenção é "infinita"; em outras vezes que é "por período indeterminado", devemos escolher porém, apenas uma forma de resposta. Isto por uma vez exige uma normalização das respostas aos quesitos de levantamento de dados.

A responsabilidade pelo acerto das informações anotadas deve ser compartilhada entre o usuário que informa e o pessoal da CSI que anota. Deve ser registrado juntamente com cada sistema o nome do responsável pela informação (em geral também responsável pelo sistema) designado pelo chefe do órgão usuário.

Uma vez digitadas as informações, a CSI disporá delas como um primeiro diretório de dados, sob a forma de um arquivo especial.

4 Usos de Diretórios de Dados.

Nas seções 1 e 2 já ficaram plenamente justificados os Diretórios de dados como elementos de um SGBD formal. Nesta seção nos restringiremos à comprovação da utilidade dos Diretórios de dados no caso de arquivos dispersos, sem a existência de um SGBD propriamente dito, onde os diretórios foram formados a partir das informações obtidas no levantamento discutido no item anterior.

A primeira utilidade é:

- permitir que todos os usuários saibam que informações estão disponíveis na empresa e onde encontrá-las

Para isso é preciso publicar o diretório de dados, o que envolve principalmente as duas listagens seguintes:

- listagem segundo a hierarquia já mencionada usuário/sistema/arquivo/item
- listagem remissiva por item, indicando para cada item a sua definição e onde ele aparece (em que arquivo de qual sistema sob a guarda de que usuário)

Para a listagem remissiva é imprescindível ter as identificações padrão de item referidas anteriormente. Pode-se também fornecer uma listagem remissiva indireta, fornecendo

para a identificação de ítem própria de cada usuário a identificação padrão correspondente, a partir da qual poderá ser consultada a listagem remissiva comum por identificação padrão.

Outra utilidade do diretório de dados é:

- indicar possíveis racionalizações sobre os arquivos existentes. As listagens remissivas evidenciarão a existência de duplicações, ou seja, dois ou mais arquivos contendo total ou parcialmente a mesma informação. No nosso exemplo (ver rodapé no início ítem 3), se para seu sistema de recursos humanos o gerente de treinamento necessita de informações sobre as habilitações requeridas por tarefa, ele provavelmente terá criado um arquivo que duplica o arquivo REQ do gerente de engenharia.

A existência de duplicações não é necessariamente um mal em si, mas precisa ser conhecida. Se não é prático para o gerente de treinamento usar o arquivo REQ diretamente (talvez por estar em local diferente e ser precário ou inconveniente o acesso por teleprocessamento) então algum procedimento deve ser adotado para que seu arquivo mantenha os dados como cópia tão atualizada quanto possível dos dados contidos no arquivo REQ.

Em certos casos, a CSI em entendimento com os usuários, pode concluir que é possível uma fusão parcial ou total dos arquivos, ou a comparação periódica de arquivos. Por exemplo, poderia ser conveniente que os arquivos ATR e EMP fossem cotados periodicamente para garantir que os empregados em ATR estão todos também em EMP e para determinar que empregados não estão no momento ligados a nenhum projeto (empregados em EMP mas não em ATR).

As tentativas de racionalização podem levar ou não a adotar-se um SGBD formal. Nem sempre é o caso de implantar um SGBD em vista dos custos envolvidos, da necessidade de pessoal treinado etc.

Pode-se concluir ser conveniente a adoção futura de um SGBD e começar então a desenvolver providências preliminares. É nesse ponto que convém levantar algumas características de organização física dos arquivos. Por exemplo, pode ocorrer que o ítem de nome do empregado contenha os nomes por extenso em EMP, mas apenas o sobrenome em CHF (como "Miranda" em nosso exemplo), sendo então o caso de procurar unificar a representação, inclusive em termos de representação interna dos caracteres (códigos ASCII ou EBCDIC, por exemplo).

O levantamento de estatísticas de uso e a pesquisa sobre necessidades não satisfeitas (ou não satisfeitas em tempo útil) de informação podem fornecer alguns dos fundamentos para decidir se a implantação de um SGBD é de fato oportuna.

Uma terceira utilidade do diretório de dados é:

- justificar a necessidade de novos sistemas e mostrar a interação destes com os sistemas existentes

Essa terceira utilidade se relaciona com um dos principais dilemas que uma CSI costuma enfrentar ao ser criada: seu objetivo final poderia ser substituir os diversos sistemas existentes, no menor tempo possível, por um grande sistema integrado sem incoerências e com um mínimo de duplicações. Acontece que isso é uma tarefa de longuíssimo prazo, e enquanto isso a empresa não pode parar, sendo necessária em alguns casos a criação de novos sistemas pelos diversos usuários. O resultado prático acaba sendo frustrante.

Como julgar objetivamente se esses sistemas propostos são necessários realmente, que forma deverão tomar e como interagirão com os sistemas existentes? Um diretório de dados pode ajudar a responder a essas perguntas, porque indica quais são as informações que os sistemas existentes já produzem. Quem propõe um novo sistema tem então a obrigação de mostrar que não estará produzindo informação já disponível e, quando for o caso, que a informação que produzirá poderá ser utilizada eficazmente por sistemas existentes.

Uma quarta utilidade é a determinação clara das responsabilidades sobre a atualidade, correção e precisão dos dados, indicando origens dos dados e mecanismos de controle. Desta forma pode-se manter em cheque tanto as tentativas de multiplicação de dados bem como as inconsistências e incorreções que poderão infiltrar-se nos arquivos.

Finalmente uma quinta utilidade é a determinação dos critérios de validação dos dados, desta forma assegurando que os dados nos arquivos foram transcritos corretamente e satisfazem os critérios de validade. Além disso tem-se agora a certeza que itens com cópias múltiplas têm exatamente o mesmo formato e representação (tipo) interno.

5 Manutenção de Diretórios de Dados.

Para que um diretório de dados retenha sua utilidade é preciso que esteja sempre atualizado.

A manutenção incide tanto em épocas em que novas aplicações forem desenvolvidas, quanto em épocas em que aplicações já existentes forem alteradas para atender a novas necessidades. Em ambos os casos é importante haver uma interação constante com todos aqueles que poderão desenvolver e/ou modificar sistemas de aplicação.

Para garantir a correção de programas o diretório deve possuir indicações de uso dos itens de dados nos diversos programas. Assim, se houver alguma alteração na descrição de um dado, pode-se saber em pouco tempo quais os programas afetados e ajustá-los. Este ajuste pode, em muitos casos ser automatizado, pois consiste meramente na transcrição de dados descritivos para os programas afetados.

6 Requisitos funcionais do Diretório de Dados.

A seguir descreveremos as condições que devem ser satisfeitas pelo diretório de dados. Não nos preocupamos nesta seção em evitar a repetição de justificativas mencionadas anteriormente neste texto.

- a. Especificação clara dos dados - para poderem especificar fluxos de informação, os projetistas de sistemas têm que ser capazes de analisar de maneira precisa os dados e fluxos de dados existentes. A ênfase deve ser em especificar e descrever dados de tal maneira que fique evidente, para qualquer pessoa interessada, como utilizar um determinado dado.
- b. Acesso simples e facilitado - os utilizadores de dados devem ser capazes de selecionar precisamente aqueles itens de dados descritivos (meta-dados) constantes do diretório que são de interesse. Esse requisito implica na possibilidade de acesso a meta-dados através de chaves padronizadas e através de pesquisa associativa (por assunto). Isto implica, em última análise, na existência de tabelas de referências cruzadas.
- c. Capacidade de determinar a existência de definições inconsistentes e/ou redundantes - a existência de dados idênticos ou muito semelhantes em diversas partes de uma instituição, são motivos para preocupação. Em alguns casos deseja-se manter redundâncias explicitamente por razões técnicas. Nestes casos deve-se proceder a uma revisão periódica das justificativas de necessidade.

dessas redundâncias. Redundâncias involuntárias devem ser evitadas a todo custo. O sistema de manutenção do diretório de dados deve providenciar auxílios para que redundâncias involuntárias sejam facilmente detetáveis.

- d. Localização de dados - se um analista não souber da existência ou da localização de um determinado dado é altamente provável que seja induzida alguma redundância. Deve-se então ter meios de saber, não só a definição, mas também onde se encontra o dado e como ter acesso a ele.
- e. determinação dos usuários dos dados - é perfeitamente possível que diversas seções de uma instituição compartilhem da utilização de um (ou mais de um) mesmo dado. Uma dessas seções será a responsável por este dado. Se as seções que compartilham do uso de um dado não forem explicitamente determinadas, poderão ocorrer transtornos em outras seções caso esse dado seja eliminado ou atualizado sem aviso pela seção responsável por ele.
- f. determinação do responsável pela integridade e especificação dos dados - apesar do administrador de dados ser a pessoa responsável pela integridade do banco de dados, ele precisa delegar poderes aos usuários. É necessário então estar bem definido quais são os direitos e deveres destes usuários com relação a cada dado.
- g. suporte ao sistema de gerência do banco de dados - de maneira geral o software que provê facilidades de bancos de dados possui os seus módulos descritores próprios. Deve ser prevista, então, uma forma para gerar estes módulos a partir do diretório de dados.
- h. suporte à validação dos dados - quase todas as instituições requerem procedimentos de validação de dados razoavelmente complexos. Estes procedimentos devem estar descritos, ou pelo menos especificados, no diretório de dados. Esta descrição poderá incorporar procedimentos programados para efetuar a validação. Desta forma a geração de um programa de validação resumir-se-ia a uma mera transcrição (automática) dos trechos de código de validação.
- i. expansibilidade - tanto os sistemas de suporte quanto as aplicações evoluem no tempo. Esta evolução deve ser necessariamente acompanhada pelo diretório de dados. É imperioso, portanto, ele ser alterável com facilidade para incluir novas descrições e para permitir a substituição de porções descritivas tornadas obsoletas.

7 Conclusões.

A existência de um diretório de dados é em parte um fator de centralização na empresa. O diretório de dados centraliza a descrição sobre o banco de dados. Nada impede, porém, que o "banco de dados" seja um conjunto de arquivos não interligados e, até, distribuído fisicamente sobre várias instalações. Portanto, o diretório de dados não impõe a centralização na guarda das informações, apenas permite determinar centralmente que informações existem e padronizar nomes e descrições de dados.

Assim a existência de um diretório de dados não deve ser temida como prejudicial à flexibilidade dos vários órgãos da empresa para dispor das informações que precisam para executar com eficiência suas incumbências. Note-se inclusive que a tendência atual é em favor de bancos de dados distribuídos, de modo que mesmo adotando um SGBD a empresa não estaria necessariamente centralizando a guarda dos dados.

No caso de banco de dados distribuídos torna-se mais importante ainda a existência de um diretório de dados, pois, de outra forma, o sistema distribuído tornar-se-ia inconsistente após algum tempo.

O conceito de diretório de dados não se prende apenas a haver ou não um SGBD implantado. Pode ir além dos arquivos computarizados existentes, registrando a existência de arquivos sob a forma de documentos, microfichas, etc.

Visto dessa forma ampla, o diretório de dados é o principal instrumento para o desempenho da função de administração de dados, [Newel,73].

Agradecimentos

Os autores agradecem ao Prof. C. J. P. Lucena por ter sugerido este trabalho, ao comandante J. da Cunha Faria por ter criado as condições para a experimentação prática de idéias aqui desenvolvidas.

Referências Bibliográficas:

- Plagman, B.K.; Altshuler, G.P.
 'A Data Dictionary/Directory System within the Context of an Integrated Corporate Data Base'
American Federation for Information Processing Societies (AFIPS) Fall Joint Computer Conference (FJCC); Volume 41, parte II; 1972; pags 1133-1140
- Uthrowczik, P.P.
 'Data Dictionary/Directories'
IBM Systems Journal 12(4); 1973; pags 332-350
- Fox, C.O.
O Administrador de Dados
 Tese de Mestrado, Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro;
 Agosto 1977
- Tsichritzis, D; Klug, A.
The ANSI/X3/SPARC DBMS Framework
 Technical Note 12; Departament of Computer Science, Universidade de Toronto; 1977
- Canning Publications
The Cautious Path to a Data Base
EDP Analyzer 11(6) Junho 1973.
- Bontempo, C.J.
 'Data Resource Management'
Data Management 11(2); Fevereiro 1973, pags 31-37
- Newel, R.A.
 'Data Administration as the Nerve Center of a Company's Computer Activity'
Data Management 11(10); outubro 1973; pags 26-31
- SIGMOD
 'The British Computer Society Data Dictionary Systems Working Party Report'
Data Base 9(2); Dezembro 1977 e
SIGMOD Record 9(4); Dezembro 1977

Apêndice A

"Lay out" do Diretório de Dados

Identificação

Nome por extenso
 Definição do item
 Nome de código
 Nome para relatório de saída

Características físicas

Valores permissíveis
 Código de validação
 Formato de representação interna
 Formato de representação de entrada
 Codificação
 Formato de edição de saída

Responsabilidades

Responsável pela aquisição
 Responsável pela manutenção
 Procedimentos de aquisição
 Formulários de entrada
 Frequências de atualização ou criação

Segurança

Criticalidade quanto à precisão
 Criticalidade quanto ao acesso indevido
 Criticalidade quanto à destruição acidental ou
 proposital
 Direitos de acesso
 Direitos de atualização
 Direitos de eliminação

Utilização

Quais os arquivos que contêm o dado
 Quais os programas que utilizam o dado
 Quais os programas que geram o dado
 Saídas que necessitam do dado
 Saídas que contêm transcrição diretas do dado
 Quem necessita do dado, direta ou indiretamente
 Frequência de utilização
 Quais os procedimentos não mecanizados que utilizam o dado

Relacionamento

Em que registros ocorre

Quais os outros dados nestes registros

Quais os dados que constituem agregados

Quais as relações a serem mantidas com respeito a outros dados

Quais as operações sobre os dados

Quais as condições das operações sobre os dados

Custo

Estimativa de valor do dado correto

Estimativa de valor (ou custo) do dado incorreto

Estimativa do custo de inexistência do dado

Estimativa do custo de aquisição

Estimativa do custo de manutenção

DEPARTAMENTO DE INFORMÁTICA

Pontifícia Universidade Católica do Rio de Janeiro