

Fèlix Saltor

Facultat d'Informàtica de la U.P.B.

Barcelona (Catalunya)

RESUM

S'estudia l'evolució dels Diccionaris de Dades, des de ser - considerats una aplicació independent i marginal per a l'admi- nistrador de (bases de) dades, passant per la seva col.labora- ció amb el Sistema de Gestió de Bases de Dades (SGBD), fins a esdevenir el cor mateix del SGBD, com es preveu en les arqui- tectures que segueixin el model ANSI/SPARC, i s'hi fan les - consideracions corresponents.

ADMINISTRACIÓ DE LES "DADES SOBRE LES DADES"

És ben sabut que l'òptica de considerar les dades d'una orga- nització com a un recurs propi de tota l'organització les fa emmagatzemar d'una manera estructurada, no lligada a cap apli- cació en particular, sinó oberta a llur utilització per totes les aplicacions (actuals i futures) que ho necessitin i amb - uns mecanismes d'integritat que les protegeixin i de segure-- tat enfront d'usos no autoritzats, és a dir, com a Base de Da- des (BD) gestionada per un Sistema de Gestió de Bases de Da-- des (SGBD) i sota el control de la funció anomenada Adminis-- tració de Bases de Dades (ABD). Ara bé, què passa amb les "da- des sobre les dades"? Els és aplicable aquesta òptica? Si s'em- magatzemen com a BD, han de tenir estructuracions específi- - ques? i interaccions privilegiades amb el SGBD? D'aquestes i altres qüestions, vegem-ne l'evolució.

Dades sobre les dades

En primer lloc, quines són aquestes "dades sobre les dades", també anomenades meta-dades? Tal com dèiem a (Saltor 76), cal diferenciar les dades concretes de les dades-patró, i els atributs de les dades concretes (valor actual; moment i causant de la darrera modificació; etc.) dels atributs de cada dada-patró que són els que donen lloc a les "dades sobre les dades" (en el món de les representacions): conjunt de valors possibles (o "tipus") per a les dades concretes d'aquesta dada-patró -com a llista, com a subconjunt d'un conjunt més gran ja conegut, o com a rutina de verificació- i en particular la "mena" de la dada: alfabètic, numèric, boo- leà, etc, i la longitud -única si fixa, mínima i màxima si va- riable-; unitat de mesura si la dada és numèrica; significa- ció dels codis si està codificada; format (s) extern (s) per

a impressió/visualització; especificació del (s) pany (s) d'accés i d'altres mecanismes de seguretat amb que el SGBD (si les dades concretes són en BD) i/o el sistema operatiu protegeixen les dades concretes d'aquesta dada-patró; usuari responsable de les dades concretes; en el cas d'incloure dades sobre la representació de les dades concretes, i si aquestes estan informatitzades, el format intern i la posició relativa o el codi distintiu de la dada concreta dins l'agrupació (concreta) superior de dades; quan es tracta d'agrupacions de dades, quins són els seus components i quins predicats d'integritat han de satisfer; quan es treballa amb "tipus abstractes de dades", independents de la representació, les operacions de les seves definicions; etc.

### Diccionari de Dades

Com que cada dada-patró rep un nom que l'identifica unívocament (sovint rep a més altres noms : noms sinònims alternatius, més entenedors; noms específics per a un llenguatge de programació o un SGBD; noms d'impressió per a capçaleres d'informes), és lògic que el lloc on s'emmagatzemen les dades sobre les dades-patró s'anomeni Diccionari (Diccionari de dades) (DD): dóna el significat de cada un d'aquests noms, tant més si conté una definició textual precisa de cada dada-patró (sovint redactada per l'usuari responsable), a més d'altres descripcions més formalitzades (com el designador o descriptor en "of language" (Muehl 78).

S'aprofita a més a més el Diccionari per a enregistrar-hi informació sobre les entitats que utilitzen les dades concretes de cada dada-patró : quins programes les llegeixen, les actualitzen; quines transaccions i feines (informatitzades o no); quins terminals; quins usuaris; quines aplicacions; -si són en BD- sota quines "visions externes" o esquemes externs apareix i com, etc.

I tot això cal considerar-ho en el marc de l'evolució en el temps de les meta-dades. Durant la fase de disseny infològic (o anàlisi funcional) no es tindran totes les meta-dades de les dades noves; a la fase de disseny tecnològic (anàlisi orgànica) es completaran aquelles i s'afegiran entitats noves -agrupacions de dades, programes, feines, etc.- i interrelacions; a la fase de proves podran introduir-s'hi canvis. Interessa doncs enregistrar amb cada meta-dada a quina fase de desenvolupament correspon. I tenint en compte que els dissenys nous poden afectar el que ja és en fase d'exploració, -i de

realització-, cal preveure que tota entitat pugui tenir jocs de meta-dades diferents per a cada fase -i fins i tot més d'un per a alguna-.

### Administració de Dades

Si l'organització considera valuoses aquestes dades sobre les dades-patró cal que estableixi la funció generalment anomenada Administració de Dades (AD), que recollirà i mantindrà aquestes meta-dades en el DD, i les explotará segons calgui. (Per a una descripció més àmplia de la figura d'AD, vegeu (Campderrich 78). Això podria fer-se manualment, però es comprén que es tasca adequada per a ser informatitzada. Tindren doncs uns programes de tractament del DD, que, junt amb una estructuració determinada de les meta-dades en el DD, formaran el Sistema de Diccionari de Dades o SDD. Aquest SDD és en cara, en aquesta primera concepció, una aplicació més, una eina de l'AD.

### AD abans que BD

Així doncs, després de l'etapa "inicial" de la informatització, en que aquestes dades sobre les dades no es recullen o figuren en part disperses en la documentació de programes i aplicacions, vé una segona etapa (al llindar de la qual es troben tot just algunes organitzacions de casa nostra) en que sorgeix una aplicació nova, interna al Servei d'informàtica, i que gestiona centralitzadament aquestes meta-dades. Aquesta segona etapa metodològicament hauria de precedir l'adopció de l'òptica de BD o ser-ne el primer pas : la decisió d'anar a BD hauria de suposar la funció d'AD, i, per a cada dada-patró, la recollida i enregistrament en el DD de les seves meta-dades (o almenys de les principals) abans de l'emmagatzematge de les dades concretes en la BD. Però usualment passa a l'inrevés : S'implanta l'òptica BD sense AD, i és més tard, quan la integració de les aplicacions per a satisfer més necessitats d'informació (que és una de les raons per anar a BD) con solida fitxers mitjançant dades suposadament comunes i fa que unes aplicacions usin dades que fins ara eren d'una altra, sense conèixer-ne les definicions exactes (el "preu de cost" del departament A no és el que el B entén per "preu de cost"; la "data de venciment" d'un usuari no és la d'un altre; etc.) quan van sortint problemes de definició de dades i s'acaba creant la funció d'AD (o ampliant la d'ABD amb funcions d'AD)

## SDD i SGBD

A l'hora de decidir com s'emmagatzemaran les dades sobre les dades en el DD cal triar si fer-ho com a BD o no. Totes les raons que, per a les altres aplicacions, o sigui per a les dades no meta-dades (o "dades per se"), juguen a favor de la BD, són també aplicables en aquest cas. Sembla doncs preferible un SDD que aprofiti les possibilitats del SGBD que s'hagi triat. En particular, com que els programes del SDD (adquirit o fet a casa) no resoldran mai totes les necessitats d'informació, caldrà un processador de llenguatge de consultes, i serà preferible que sigui el mateix que fem servir per a les BD. En el passat s'han triat SDD independents de SGBD, degut a l'absència o insuficiència de SDDs amb DD en forma de BD, però avui en dia per a la majoria de SGBDs existeixen SDD que aprofiten les seves possibilitats.

En un i altre cas és clar que el SDD és una aplicació més (interna al servei d'informàtica), una eina de l'AD marginal per al SGBD.

Si és convenient, com dèiem abans, que el DD precedeixi la BD, pot semblar que la decisió sobre SDD pot ser anterior a la decisió sobre quin SGBD emprar, i per tant triar un SDD independent, no condicionat per cap SGBD donat. Però no fora lògic que tingués més pes el temps en que tindrem DD sense SGBD, en front del temps més llarg en que coexistiran tots dos, i d'altra banda pot anar-se establint la metodologia de recollida de les meta-dades sense tenir el SDD, i decidir conjuntament sobre SDD i SGBD.

Quan es disposa, o s'ha decidit disposar, de més d'un SGBD (en una o múltiples CPUs, en una o múltiples localitats), la decisió sobre SDD és més difícil. Per una banda, convé que el DD estigui en forma de BD; i pot triarse aquell dels SGBDs que es fan servir que resulti més escaient, però per altra és possible que només un SDD independent tingui ja previstes per a tots els nostres SGBDs les interrelacions de que ara parlarem

## INTERFÍCIES ENTRE SDD I SGBD

### Interfícies "transparentes" al SGBD

El SGBD necessita per a funcionar (i per a facilitar la inde-

pendència dades-programes que és una de les seves característiques) unes descripcions dels formats de les dades concretes de totes les dades-patró (elementals i agrupacions) de la BD i dels mètodes per al seu accés i actualització, que se li donen en un llenguatge de descripció de dades (DDL) i que es compilen i enregistren internament d'una manera específica per a cada SGBD, i que s'anomenen Esquema en la terminologia CODASYL-DBTG, Data Model Definition en la relacional, Data Base Description en la DL/I, etc., i que anomenarem Esquema intern seguint a ANSI/SPARC. També necessita de les descripcions que els programes i usuaris han de tenir de les dades de la BD : Són els Sub-Esquemes (CODASYL-DBTG), Data Sub Model Definitions (model relacional), PSBs (DL/I), o Esquemes externs. Totes aquestes descripcions figuren també lògicament al DD, i convé doncs evitar llur redundància externa : que paral·lelament i separatament l'AD les gestioni al DD mentre l'ABD ho fa al SGBD. Com que les meta-dades del DD són més àmplies que aquestes meta-dades del SGBD, sembla preferible que la gestió es faci sobre el DD, i sigui a partir d'aquest, -a base que el SDD pugui produir les descripcions en els DDLs adequats-, que s'actualitzin els Esquemes del SGBD (o SGBDs). S'acostuma també a unificar les funcions d'AD i ABD en una sola, excepte en organitzacions grans.

Anàlogament pel que fa a les especificacions de seguretat: que el SDD les passi al SGBD en el format que aquest necessiti.

Els programes i especificacions de consultes o informes necessiten també d'unes descripcions de les dades amb què treballen (més reduïdes i independents dels formats interns que en l'òptica pre-BD) en la sintaxi corresponent a llurs llenguatges. També aquí s'evitarà la redundància externa de preparar-les manualment si el SDD pot produir-les automàticament per als traductors (compiladors, assembladors i intèrprets) dels llenguatges de programació/consultes/informes de la instal·lació.

(Els SDD existents permeten generalment també les interfícies corresponents en sentit contrari: per a emmagatzemar en el DD les meta-dades a partir dels esquemes del SGBD i de les descripcions de dades dels traductors. Això es útil en una implantació del DD posterior a les BD, però no és aconsellable en una explotació normal de DD i BD).

El SDD pot també produir mòduls de verificació de dades d'entrada, en els diferents llenguatges, a partir de les meta-dades del DD, així com jocs de prova. Més en general, s'està in

vestigant sobre la producció automàtica de software a partir del DD.

Si posem en obra mecanismes (p. ex. del sistema operatiu) per a impedir que les actualitzacions dels Esquemes del SGBD i de les descripcions de dades dels programes (i especificacions de consultes o informes) puguin venir de font altra que el DD, i instituïm doncs aquest DD en el repositori central únic de les meta-dades, tindrem una segona concepció del DD, ja no com una aplicació més sino com una aplicació central.

### Interfícies conegudes pel SGBD

Així com aquesta segona concepció ampliava la primera, establint una col.laboració unidireccional entre SDD i SGBD/traductors, però "transparent" per a aquests (en el sentit de poder "ignorar" la presència del SDD), és lògic pensar en una tercera concepció del SDD en que les col.laboracions siguin recíproques. Això pot fer-se introduint en el SGBD i els traductors rutines (en principi opcionals) que ofereixin interfícies amb el SDD. Aquesta col.laboració permet millores diverses:

- Com a resultat de cada execució d'un traductor, enregistrar, automàticament en el DD les interrelacions del programa compilat/assemblat amb les seves visions de les dades.
- que les estadístiques d'accesos a BD recopilats pel SGBD siguin emmagatzemades al DD.
- que el SGBD comprovi, abans d'executar cap programa, que tant ell com el seu esquema extern corresponen a una fase de desenvolupament (proves, explotació) compatible amb la fase de la BD amb què vol treballar.
- que el SDD passi els esquemes al SGBD no en DDL, que necessita compilació, sino directament en forma objecte; o bé que el SGBD els vagi a buscar al SDD (i anàlogament pel que fa als traductors).
- anàlogament respecte a les especificacions de seguretat.
- que, en el cas d'un SGBD integrat amb un gestor d'aplicacions en teleprocés (anomenat sovint sistema DB/DC), aquest, en l'engegada, tregui del DD la configuració de Bases de Dades/Programes/Xarxa de teleprocés que se li indiqui.

## INTEGRACIÓ DEL SDD DINS DEL SGBD

El darrer pas en l'augment de la col.laboració entre SDD i SGBD és la integració dels dos. Amb ella ja no hi haurà un DD per un costat i una biblioteca d'esquemes del SGBD per un altre, sino un repositori únic (que podrà tenir les redundàncies internes que l'eficiència imposi), sobre el qual aquest SGBD integrat farà directament totes les comprovacions i actualitzacions, i en particular les comprovacions de seguretat, d'integritat (tant de valors correctes de dades elementals, com dels predicats de les agrupacions de dades), de compatibilitat de fase, etc.

El cas de les dades no gestionades pel SGBD, és a dir, les dels fitxers, posarà una dificultat a aquesta integració si el SGBD integrat no ho preveu. Però es poden integrar les altres dades, mantenint un DD separat per als fitxers (solució transitòria), o bé que els fitxers hi siguin definits com a BD, com molts SGBDs ja permeten, i tenir una integració total.

En aquesta concepció, el DD esdevé el cor del SGBD, i es pot anomenar "Data Dictionary/Directory" (DD/D) o "Dictionnaire/Repertoire de Données" (noms que alguns ja donen als DD de la que hem anomenat segona concepció), o bé, tal com proposa Nijssen, "Information Dictionary" o Diccionari de la Informació (Gradwell 78) arriba a proposar "Enciclopèdia de Sistemes").

L'arquitectura marc per a SGBDs proposada pel grup d'estudi sobre SGBDs del comité SPARC d'ANSI/X3 -vegeu-la a Tschritzis & Klug (77), o bé a Hernán (78)-, i que tant està influïnt i influirà en tots els desenvolupaments en aquesta àrea, col.loca precisament un DD/D en el centre del SGBD. (Aquesta arquitectura marc contempla tres nivells d'esquemes: intern, conceptual i externs; i un Administrador de l'esquema conceptual -"enterprise administrator"- que correspon en gran part al que hem anomenat AD). I és notable que el posar el DD/D al cor del SGBD no hagi estat el punt de partida d'aquest grup de treball d'ANSI/SPARC, sino que aquests, en el disseny de l'arquitectura marc, han anat trobant la necessitat de repositoris per a diverses meta-dades, i finalment s'han adonat de la conveniència d'integrar-los en un únic DD/D, nucli central del sistema.

Altres autors (Hurtubise 76 i 77) i altres esdeveniments situen també el DD/D com a centre de control del SGBD integrat.

La direcció en la qual es va sembla doncs ser aquesta de la integració del DD al cor del SGBD; podem preparar-nos-hi ja des d'ara.

#### NOTES

1.- Aquesta col.laboració progressiva entre SDD i SGBD que s'ha descrit és independent i paral.lela a la col.laboració progressiva entre SGBD i Sistema operatiu.

2.- Agraeixo els comentaris que a una versió anterior d'aquest escrit m'han fet J. Guinovart, B. Camderrich, J.L. Solanas i J. Farrús.

#### BIBLIOGRAFIA

BCS.DDSWP (77)

"The British Computer Society Data Dictionary Systems Working Party Report" BCS, Londres o SIGMOD Record, 9, 4, ACM, NY, U USA, o SIGBDP, 9, 2, ACM, NY, USA.

R. Canning (74)

"The Data Dictionary/Directory Function" EDP Analyzer, 12, 11. Vista, Californica, USA.

R. Canning (78)

"Installing a Data Dictionary" EDP Analyzer, 16, 1. Vista, California, USA.

R. Hurtubise (76)

"Informatique et Information" Agence d'Arc, Montréal, Québec, Canadà & Eds. d'Organisation, Paris, França.

R. Hurtubise (77)

"La gestion de l'information" Agence d'Arc, Montréal, Québec, Canadà & Eds. d'Organisation, Paris, França.

H. Lefkovitz (77)

"Data Dictionary Systems". QED Information Sciences Inc. Wellesly, Massachusetts, USA.

JR. Muehl (78)

"DB/DC Data Dictionary Release 2 Implementation Guide" IBM (G320-6017). Palo Alto, USA.



F. Saltor (76)

"Dels fitxers clàssics a les bases de dades : Dades i models estructurals". Novàtica, nº 11. Barcelona

D. Tsichritzis & A. Klug (77)

"The ANSI/X3/SPARC DBMS framework". Technical Note 12, CSRG, University of Toronto, Toronto, Ontario, Canadà.

P. Uhrowczik (73)

"Data dictionary/directories" IBM Systems Journal, 12, 4, p 332. Armonk, NY, USA.

J. Hernán (78)

"El modelo de arquitectura de sistemas de bases de datos según ANSI/SPARC" Novàtica, nº 24. Barcelona.

B. Campderrich (78)

"La figura de l'Administrador de Dades" Novàtica, nº 24. Barcelona.

D.J. Gradwell (78)

"Objectives and scope of a data dictionary system" The Computer Bulletin, des. 1978. Londres.