

DATA MART ȘI OLAP ÎN ANALIZA ȘI EVALUAREA POLUĂRII APELOR INDUSTRIALE

Constanța Zoie Rădulescu

Institutul Național de Cercetare - Dezvoltare în Informatică, ICI, București

e-mail radulescu@ici.ro

Rezumat: Lucrarea prezintă un mod de rezolvare a unor probleme legate de deciziile privind poluarea apei în industria textilă, în contextul unei dezvoltări durabile. Se prezintă, astfel, arhitectura unui sistem suport de decizie, bazat pe modele ale analizei multicriteriale, alături de un model multidimensional al datelor. Cu ajutorul unei colecții de date specifice (data mart) și a modelului multidimensional al datelor, se poate realiza o analiză de tip prelucrare analitică on – line („On line Analytical Processing” - OLAP).

Se descrie, mai întâi, contextul în care s-a pus problema definirii unui astfel de sistem. Se prezintă, astfel, conceptul de dezvoltare durabilă, precum și problemele privind mediul și poluarea apei, specifice industriei textile, în contextul integrării României în Uniunea Europeană. Se descriu, apoi, conceptele generale, necesare definirii arhitecturii sistemului (sistem suport de decizie, magazie de date, data marts, prelucrare analitică on line – OLAP). Luând în considerare aceste concepte, se prezintă etapele necesare în procesul de rezolvare și găsire a unei soluții de compromis pentru două probleme importante: reducerea nivelului de poluare și creșterea eficienței economice. Aceste probleme sunt două dintre problemele importante cu care se confruntă managerii unităților economice din cadrul industriilor ale căror procese tehnologice au impact asupra mediului.

În final, se descrie arhitectura sistemului suport de decizie.

Cuvinte cheie: SSD, data marts, OLAP, dezvoltare durabilă, managementul riscului.

1. Introducere

Dezvoltarea industrială din ultimul secol a determinat, printre altele, apariția de efecte negative și riscuri privitoare la cadrul social și natural în care aceasta se desfășoară. Problemele de decizie, legate de elaborarea unor strategii de dezvoltare a activităților industriale și de impactul acestora asupra mediului, au devenit din ce în ce mai complexe. Instrumentele software care pot ajuta decidentul în luarea deciziilor de bună calitate în domeniul unei dezvoltări industriale durabile sunt sistemele suport de decizie (SSD). O mare parte sunt sisteme bazate pe metode matematice privind decizia multicriterială și managementul riscului. Datele necesare proceselor de decizie au următoarele caracteristici: a) un volum mare, b) un grad redus de coerență, c) surse de proveniență diversă: fișiere, colecții de fișiere, baze de date operaționale, c) caracter istoric. Aceste caracteristici au determinat organizarea datelor în magazine de date și colecții de date specifice (data marts), organizare ce permite o analiză de tip OLAP (prelucrare analitică on line).

Printre cele mai utilizate metode de decizie, în contextul unei dezvoltări industriale durabile, sunt metodele de decizie multicriterială. De exemplu, problemele de decizie privind dezvoltarea durabilă implică o balanță între obiective conflictuale cum ar fi profiturile realizate de activitățile industriale și nivelurile de poluare a mediului.

Scopul acestei lucrări este de a prezenta un mod de rezolvare a unor probleme legate de deciziile privind poluarea apei în industria textilă în contextul unei dezvoltări durabile. Se prezintă, astfel, arhitectura unui sistem suport de decizie, bazat pe modele ale analizei multicriteriale, alături de un model multidimensional al datelor. Cu ajutorul unei colecții de date specifice (data mart) și a modelului multidimensional al datelor, se poate realiza o analiză de tip prelucrare analitică on – line („On line Analytical Processing” - OLAP).

Lucrarea este organizată după cum urmează: Capitolul 2 prezintă cadrul general al problemei poluării apei în viziunea integrării României în Uniunea Europeană și a unei dezvoltări industriale durabile. Capitolul 3 prezintă concepte generale ale teoriei deciziei (sistem suport de decizie, magazie de date, colecție de date specifice (data marts), prelucrare analitică on line – OLAP) ce stau la baza definirii sistemului suport de decizie pentru analiza și evaluarea poluării apei industriale. În capitolul 4, se descriu etapele necesare a fi rezolvate într-un proces de control al poluării apei industriale.

Capitolul 5 prezintă arhitectura unui sistem suport de decizie, bazat pe modele și analiză multidimensională a datelor - OLAP pentru analiza și evaluarea poluării apei industriale, în contextul unei dezvoltări durabile.

Lucrarea se încheie cu concluzii și cu o listă bibliografică.

2. Problema poluării apei în viziunea integrării României în Uniunea Europeană și a unei dezvoltări industriale durabile

Până acum douăzeci de ani, s-a acordat prioritate creșterii economice, fără a se ține cont deloc de necesitatea protejării mediului. Formele grave de manifestare a deteriorării componentelor mediului înconjurător au impus schimbarea acestei optici.

Cea mai larg folosită definiție pentru „dezvoltare durabilă” a fost dată în anul 1987 în raportul „Our Common Future” de World Commission on Environment and Development. *Dezvoltarea durabilă este dezvoltarea care îmbină cerințele prezentului fără a compromite capacitatea generațiilor viitoare de a-și satisface propriile nevoi* [9].

Dezvoltarea durabilă a constituit obiectivul mai multor reuniuni internaționale, de mare amploare, care au culminat cu Conferința Națiunilor Unite pentru Mediu și Dezvoltare, cunoscută și sub numele de „The Earth Summit” (Conferința mondială a pământului), care a avut loc la Rio de Janeiro în iunie 1992. Rezultatul acestei Conferințe a fost adoptarea AGENDEI XXI (Agenda Schimbării), document prin care se propun, la nivel de principii, soluții (într-un plan de acțiune coerent) la problemele critice cu care se confruntă omenirea. [1].

Este, în general, acceptat faptul că dezvoltarea durabilă poate fi definită în termenii a trei componente principale:

- ecologic durabil care presupune ca dezvoltarea să nu afecteze calitatea mediului înconjurător;
- economic durabil care presupune ca dezvoltarea să satisfacă necesitățile umane;
- social/cultural durabil care presupune ca dezvoltarea să fie acceptată social (echitabil și responsabil).

Împărțirea în trei componente separate, așa numita abordare „mozaic”, nu înseamnă că dezvoltarea durabilă poate fi obținută prin cele trei componente independent una față de cealaltă. Componentele trebuie considerate ca cele trei dimensiuni interdependente ale durabilității, ceea ce înseamnă că cele trei criterii de durabilitate trebuie să fie îndeplinite simultan pentru a se dobândi o dezvoltare durabilă. O imagine a celor trei componente și a legăturilor dintre acestea așa cum a fost concepută de Volker Teichert [8] este dată în figura 1.

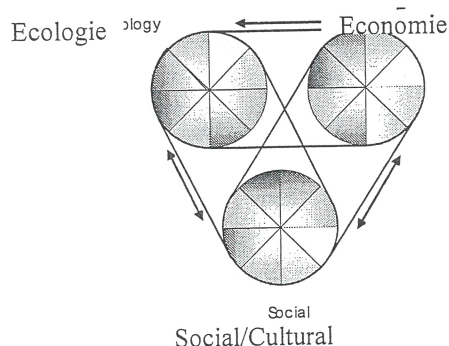


Figura 1. Componentele unei dezvoltări durabile

Starea actuală a mediului în România relevă probleme acute, în special „ca urmare a poluărilor locale, produse, în principal, în sectoarele exploatărilor petroliere și de minerit, în industriile de prelucrare a minereurilor și petrolului, termoelectrică, industria chimică, de prelucrare a lemnului și celulozei, metalurgie, siderurgie, industria electrotehnică și a construcțiilor de mașini, industria cimentului, transporturi, gospodăria comunală și agricultură” [4].

Integrarea în UE reprezintă unul dintre cele mai importante obiective ale României. Aceasta implică o muncă foarte dificilă pentru a satisface cerințele economice, sociale și de mediu ale țărilor UE. Unul dintre cele mai dificile capitole este mediul.

România a deschis negocierile pentru capitolul 22 – Mediul, în cadrul primei Conferințe de Aderare România-UE desfășurată, la Bruxelles, în perioada 21-22 martie 2002.

La 28 ianuarie 2004, Comisia europeană a adoptat „Planul Acțiunilor pentru Tehnologii de Mediu”, pe care l-a făcut public prin Comunicarea COM (2004) 38 final „Stimularea tehnologiilor pentru o dezvoltare durabilă”.

Acest plan se încadrează în problematica dezvoltării tehnologiilor de mediu (definite în Agenda 21, capitolul 34, pentru tehnologiile sănătoase) la nivel global și la nivelul Uniunii Europene.

ETAP- Planul Acțiunilor pentru Tehnologii de Mediu, a fost inclus pe agenda de lucru a Consiliului European în martie 2004. Comisia a pus în aplicare Planul și va întocmi un raport în 2006.

Pentru implementarea directivelor de mediu, în România s-au făcut eforturi susținute, iar concluziile Raportului de țară, dat publicității, pe 7 octombrie 2004, la Bruxelles, atestă aceasta, autoritățile angajându-se în finanțarea suplimentară a investițiilor.

Industria textilă este unul dintre cele mai importante sectoare economice ale României din punct de vedere al producției, forței de muncă, exportului (30% din export). În procesele textile, se utilizează o mare varietate de

substanțe chimice (enzime, detergenți, acizi, coloranți, fungicide etc.) care contribuie într-o măsură mai mare sau mai mică la poluarea mediului ambiant. În special, apele reziduale, provenite din secțiile de finisare textilă, pun serioase probleme legate de cantitatea de sedimente, pH, temperatură, culoare (grupa de coloranți), conținutul de substanțe organice (particule de fibre, materiale fibroase, tenside etc.), conținutul de substanțe anorganice (săruri, acizi, alcalii, clor, metale etc.). Datorită diversității structurii producției, calitatea apelor uzate variază nu numai de la o societate comercială la alta, ci, uneori, și în funcție de perioadă, chiar în interiorul aceleiași societăți.

O schemă simplificată privind intrările și ieșirile într-o unitate economică din industria textilă este ilustrată în figura 2.

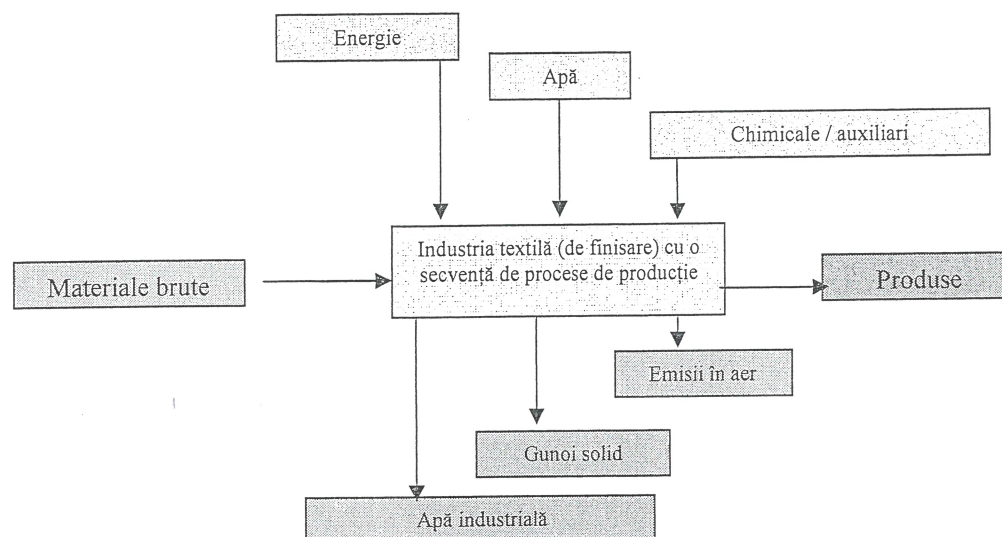


Figura 2. Intrările și ieșirile dintr-o unitate economică din cadrul industriei textile

Implementarea directivelor privind micșorarea poluării mediului, controlul poluării și managementul riscului de poluare, alături de o dezvoltare economică viabilă, pune probleme importante de decizie atât la nivel strategic, cât și la nivelul fiecărei unități economice. Pentru a se lua cele mai bune decizii, este necesară luarea în considerare a metodelor și instrumentelor teoriei deciziei. Sistemele suport de decizie, orientate pe date sau modele, devin ajutoare în luarea celor mai adecvate decizii.

3. Sisteme suport de decizie, data marts și OLAP

Vom prezenta, în continuare, câteva concepte generale din teoria deciziei, ce stau la baza definirii arhitecturii sistemului suport de decizie pentru analiza și evaluarea poluării apei industriale, provenite din procesele tehnologice din industria textilă. Vom considera, astfel, conceptele de sistem suport de decizie, magazie de date, data marts, organizare multidimensională a datelor și prelucrare analitică on line (OLAP).

O definiție din 2002 a Sistemelor suport de decizie - SSD (Decision Support Systems - DSS) este cea dată de Power [5] ca fiind „un sistem informatic interactiv, menit să-l ajute pe decident să utilizeze date, documente și modele pentru a identifica și rezolva probleme și a lua decizii”.

O definiție extinsă este dată de Filip, F.G. [2], care consideră SSD ca fiind o clasă de sisteme informatice, cu caracteristici antropocentrice, adaptive și evolutive, care integrează o serie de tehnologii informatice și de comunicații de uz general și specifice, și interacționează cu celelalte părți ale sistemului informatic global al organizației. Menirea SSD este de a atenua efectul limitelor și restricțiilor decidentului într-un număr semnificativ de activități pentru rezolvarea unei palete largi de probleme decizionale nebanale.

Două din subclassele de *tipuri generice* de SSD din cadrul conceptual propus de Power [5], preluat din Filip, F.G. [2] sunt:

1. *SSD orientate către date* („Data – driven DSS”), care se referă la acele sisteme suport pentru decizii în care subsistemul de gestiune a unor volume mari de date structurate reprezintă componenta tehnologică dominantă.
2. *SSD orientate către modele* („Model – driven DSS”), în care accentul principal se pune pe modelele matematice de simulare și optimizare (inclusiv modelele pentru analiza deciziilor multiatribut). Astfel de soluții utilizează volume mai reduse de date.

Din clasa SSD orientate către date un loc special îl ocupă magaziile de date (Data Warehouse) și prelucrarea analitică on line (On Line Analytical Processing – OLAP). Data mart este o categorie specială de magazie de date, organizată la nivelul unui departament. Ea conține un volum mai mic de date comparativ cu o magazie de date.

Inmon [3] definea o *magazie de date* ca fiind o colecție de date, destinată fundamentării deciziei manageriale, orientată pe subiect, integrată, nevolatilă și organizată pe perioade de timp (*A data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management's decision making process*).

Privind enunțul cu ochiul unui matematician, acesta ar putea fi descompus într-o definiție („o magazie de date este o colecție de date *destinate fundamentării deciziei manageriale*”) și o caracterizare („o astfel de colecție de date este (1) orientată pe subiect, (2) integrată, (3) plasată într-un context temporal și (4) permanentă (nevolatilă)”).

Fiecare dintre aceste idei joacă un rol important, în conceptul de cât de activă poate fi o magazie de date pentru a asista cerințele de management al datelor.

Spre deosebire de colecțiile de date, utilizate de sistemul operațional - orientate spre optimizarea și siguranța procesării datelor - datele dintr-o magazie de date sunt organizate într-o manieră care să permită analiza lor, deci extragerea semnificației economice pe care o poartă. Rolul unei magazii de date este de a oferi o imagine coerentă asupra datelor, relative la activitatea unei organizații și a contextului în care acesta acționează. O arhitectură a unei magazii de date și legătura cu data marts este ilustrată în figura 3.

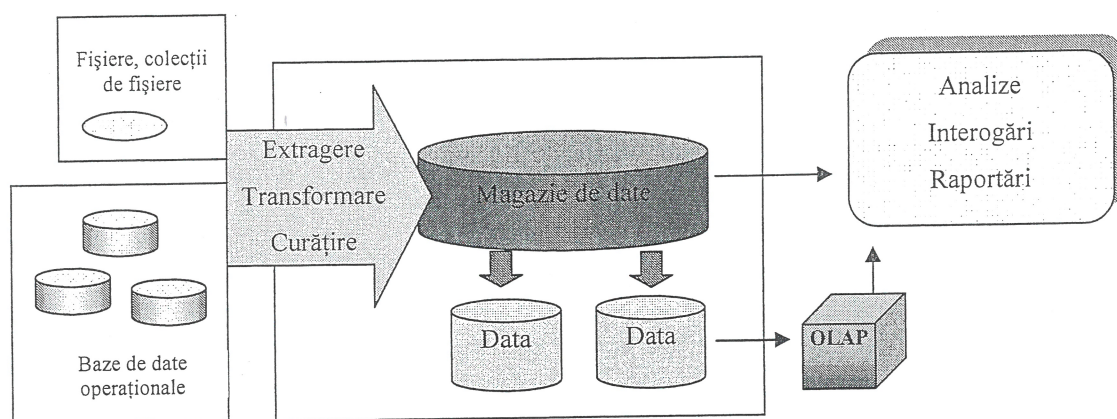


Figura 3. O arhitectură a unei magazii de date

Ca o extensie a conceptului de magazie de date, a apărut conceptul de data mart – colecție de date specifice. Acest concept a apărut ca urmare a faptului că, în cadrul unei organizații, există mai multe tipuri de analiză asupra aceleiași colecții de date. Astfel, un data mart este o magazie de date creată pentru un departament specific într-o organizație [10].

Se disting două metode principale prin care se poate valorifica informația din magazia de date: „mineritul” în date (*data mining*) și analiza multidimensională – OLAP (*On Line Analytical Processing*). Data mining este o tehnică care vizează descoperirea unor „șabloane” (*patterns*) semnificative în structura datelor, care să indice tendințe generale. Tehnicile implicate în mineritul datelor sunt tehnici complexe, de exemplu: arbori de decizie, arbori de clasificare, rețele neurale, vizualizarea datelor, logică fuzzy, statistică matematică.

Tehnicile de prelucrare analitică on line sau analiza multidimensională a datelor OLAP (*On line Analytical Processing*) permit managerilor să analizeze datele după mai multe dimensiuni și ierarhii. Unele soluții OLAP memorează date sau accesează date direct din magazia de date, altele memorează date într-un format multidimensional. Instrumentele OLAP nu sunt proiectate pentru a suporta raportarea, totuși, multe instrumente OLAP sunt integrate cu instrumente de raportare și EXCEL pentru a permite cerințe de raportare de bază.

Un exemplu de organizare tridimensională a datelor privind poluarea apei în industria textilă este ilustrat în figura 4.

Diverse vizualizări în această organizare tridimensională sunt ilustrate în figura 5.

Instrumentul OLAP permite analiza interactivă și rapidă a datelor prin operații de tip rulare (roll-up), forare (drill-down), feliere și decupare de cubulețe (slice and dice) etc. Utilizatorii pot obține rezultate imediate prin formularea de întrebări, parcurgând dinamic dimensiunile structurii multidimensionale, lucrând cu niveluri diferite de sinteză sau detalieri. Un exemplu de astfel de întrebare, pentru poluarea apei industriale ar fi „Care sunt toate măsurătorile care au valorile sub (sau peste) pragurile de risc (de alarmă) ale poluanților ?”

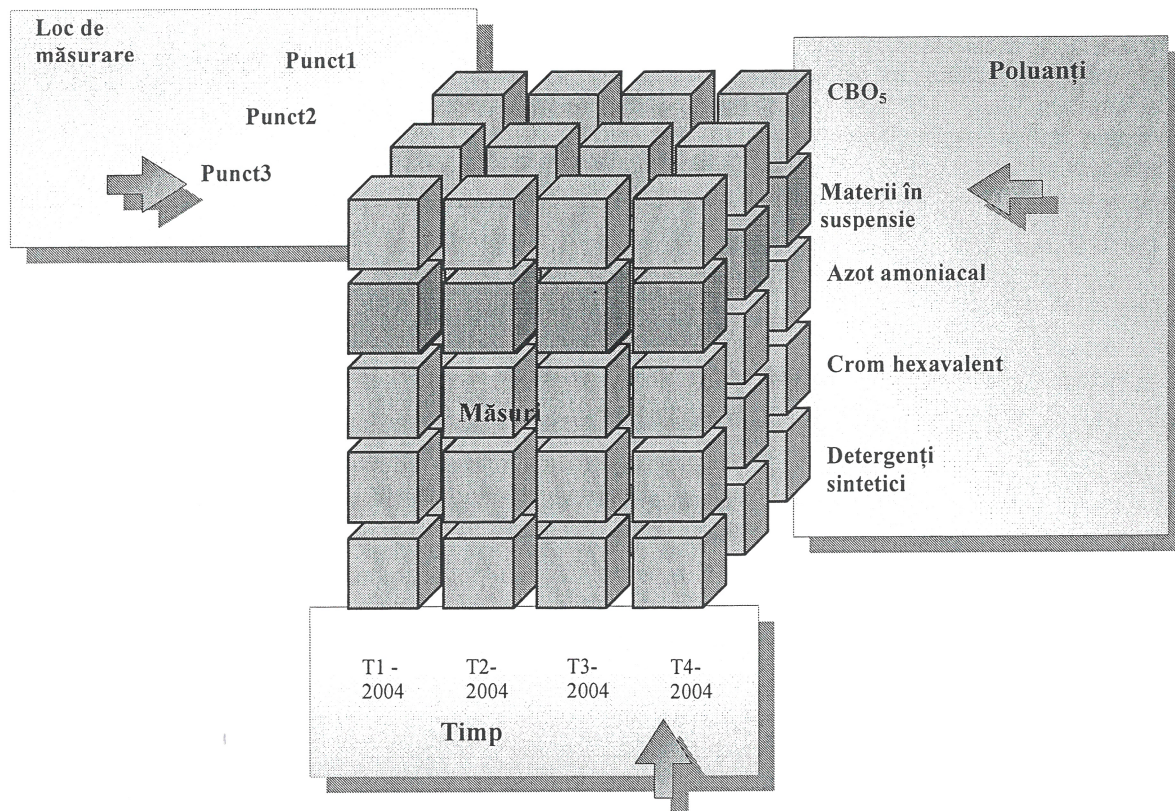


Figura 4. Organizare tridimensională a datelor privind poluarea apei în industria textilă

Vizualizarea pentru pct. de măsurare

Vizualizarea pentru poluanți

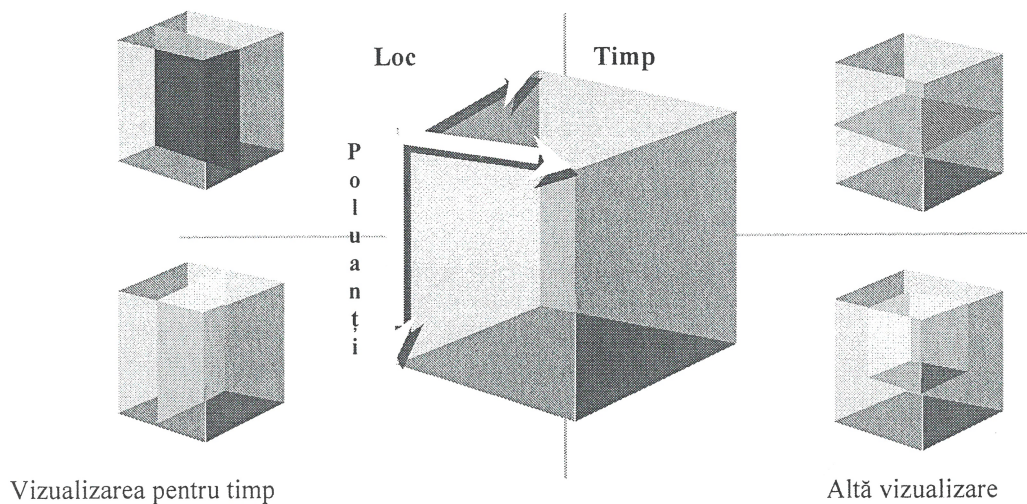


Figura 5. Diverse vizualizări în structura tridimensională

Pentru că dezvoltarea durabilă presupune în general rezolvarea unor probleme conflictuale, sistemul suport de decizie ar trebui să fie un SSD orientat către modele de decizie multicriteriale (multiatribut sau multiobiectiv). Pe de altă parte, datele capturate din sistemele informatice operaționale, existente în cadrul unităților economice (în special datele istorice) reprezintă o mină de aur din punct de vedere informațional. Aceste date pot fi gestionate, analizate și pot oferi o imagine coerentă, sintetică asupra poluării (surse de poluare, poluanți, procese tehnologice etc.). Acest lucru a determinat orientarea SSD către date și analiza lor.

Un SSD pentru controlul poluării apei industriale pentru o dezvoltare durabilă ar trebui să conțină în structura sa o componentă pentru gestionarea și analizarea datelor (de tip data mart și OLAP) și o altă componentă pentru decizia multicriterială.

4. Etape ale unui proces de control al poluării apei industriale ținând cont de o dezvoltare industrială durabilă

Managerii unităților economice din industria textilă se confruntă cu două probleme importante, reducerea poluării și creșterea eficienței economice. Pentru rezolva cele două probleme, trebuie să se găsească o soluție de compromis. Luând în considerare metodele și instrumentele teoriei deciziei, etapele necesare pentru rezolvarea celor două probleme, sunt:

1. preluarea datelor din bazele de date operaționale (categorii de poluanți, puncte de măsurare, indicatori ai poluării apei, aerului, solului pentru diferiți poluanți specifici industriei textile, costuri legate de producție, costuri legate de protecția mediului, norme de protecția mediului, praguri de alertă, praguri critice, coloranți, rețete, produse, tehnologii);
2. extragerea, transformarea, curățirea și încărcarea lor într-o colecție de date specifică (de tip data mart) unde informațiile sunt organizate pe subiecte de interes (analiza riscului poluării mediului, decizii pentru o dezvoltare industrială durabilă etc.);
3. organizarea multidimensională a datelor și folosirea instrumentelor de tip OLAP pentru evidențierea esenței datelor printr-un acces rapid, consistent și interactiv la o mare diversitate de vederi posibile ale datelor;
4. managementul riscului prin folosirea de măsuri specifice ale riscului;
5. folosirea de modele de analiză multicriterială, modele predictive și simulări de scenarii, pentru efectuarea unor analize de tipul „Ce s-ar întâmpla dacă...?” („What if...?”);
6. analiza rezultatelor.

5. Arhitectura unui sistem suport de decizie bazat pe modele și analiză multidimensională a datelor (OLAP) pentru analiza și evaluarea poluării apei industriale în contextul unei dezvoltări durabile.

Ținând cont de metodele și instrumentele pentru asistarea deciziei, prezentate în Capitolul 3, precum și de etapele prezentate în Capitolul 4, vom descrie, în continuare, arhitectura unui sistem suport de decizie, bazat pe modele (de decizie multicriterială) și analiză multidimensională a datelor (OLAP). Sistemul este destinat analizei și evaluării poluării apei industriale în contextul unei dezvoltări durabile (figura 6). El este adaptat problemelor de decizie, legate de industria textilă și, în particular, de poluarea apei datorată proceselor tehnologice. Sistemul conține un modul pentru analiza și evaluarea riscului de poluare a apei folosind o organizare a datelor de tip data mart și analiză OLAP urmărind mai multe dimensiuni ale datelor organizate în ierarhii de date.

Se încearcă prin această abordare definirea unui sistem suport de decizie, orientat atât pe date (prin includerea organizării datelor de tip data mart, o categorie specifică de magazie de date), cât și pe modele (analiza riscului, modele de decizie multicriterială).

Acest instrument este compus din patru componente principale:

- colecția de date;
- colecție de tip data mart;
- analiză multidimensională - OLAP;
- modelare multicriterială.

Componenta „Colecția de date” conține datele primare așa cum sunt ele culese din sisteme operaționale existente sau din documente primare. Se urmărește preluarea de date și informații care să poată fi integrate în modele și analize

privind poluarea apei industriale. De exemplu, componenta „Colecția de date” poate conține pe lângă datele privitoare la poluarea apei datorate unor procese chimice și date și informații privind poluanții, coloranții, rețetele folosite în diferite faze ale procesului tehnologic, faze tehnologice, efecte posibile asupra sănătății și mediului etc. Aceste date sunt completate cu informații privitoare la cadrul legal (norme și standarde) privind calitatea apei și a emisiilor de poluanți în mediu. Aceste date transformate și aduse la un numitor comun vor constitui intrările în componenta „Colecție de tip data mart”. Ele vor fi organizate pe subiecte într-o structură ierarhică.

Componenta „Colecție de tip data mart” conține trei categorii de date:

- baza de date pre-analitică;
- baza de date de decizie;
- baza de date post-analiză.

Baza de date pre-analitică cuprinde datele organizate, necesare: (a) analizelor de tip OLAP și (b) modelelor de decizie multicriteriale.

Baza de date de decizie cuprinde soluțiile obținute în urma analizelor și rezolvării modelelor. Aceste date constituie intrări pentru modelele de analiză a soluțiilor obținute (what if ?, scenarii, simulare, testare înapoi etc.)

Baza de date post-analiză conține soluțiile modelelor what if ? (ce s-ar întâmpla dacă?), scenarii, simulare, testare înapoi etc.

Componenta „Analiză multidimensională - OLAP” realizează analize după mai multe dimensiuni (poluanți, timp, loc de măsurare, proces tehnologic, tip de poluant etc.). Rezultatele analizelor sunt memorate în baza de date de decizie. Un tip special de analiză este cea pentru managementul riscului.

Mai multe informații privind analiza multidimensională a datelor OLAP, precum și un exemplu din cadrul poluării apei în industria textilă, sunt prezentate în [7].

Componenta „Modelare multicriterială” conține o bibliotecă de modele de decizie multicriterială.

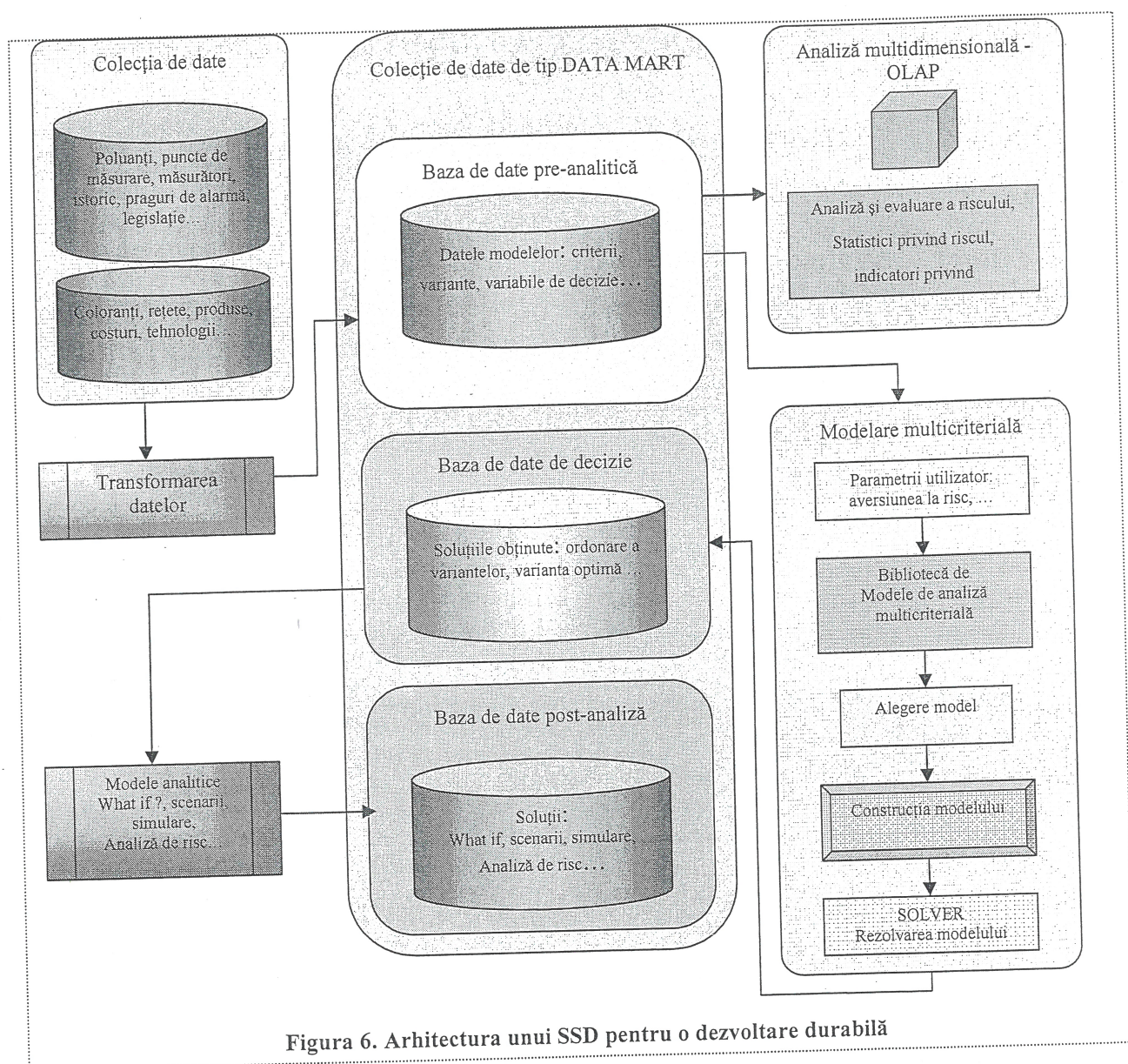
6. Concluzii

Complexitatea problemelor de dezvoltare durabilă necesită luarea de decizii raționale, iar luarea deciziei devine din ce în ce mai dificilă în special pentru probleme de mediu. Progresele realizate în cadrul teoriei deciziei și a sistemelor suport de decizie au determinat apariția de metode și instrumente ce pot asista pe decidentul de astăzi în luarea deciziilor celor mai adecvate. Totuși, pentru a ajuta în luarea de decizii pentru problemele complexe, aceste instrumente nu sunt ușor de proiectat și realizat.

Lucrarea a demonstrat, în contextul unei dezvoltări industriale durabile, necesitatea elaborării unui SSD pentru controlul poluării apelor în industria textilă bazat pe data mart, analiză OLAP și modele de decizie multicriterială.

S-a prezentat arhitectura SSD – Sistem Suport de Decizie pentru analiza și evaluarea poluării apei industriale în contextul unei dezvoltări durabile și s-au descris avantajele folosirii data mart (orientarea colecției de date către anumite subiecte (cum ar fi controlul poluării apei), integrarea și permanența datelor).

S-a prezentat modelul pentru analiza multidimensională a datelor și s-au vizualizat diverse operații OLAP. Cercetările pentru definirea arhitecturii SSD au fost realizate în cadrul etapei 1 a proiectului RELANSIN cu titlul: „Dezvoltare durabilă prin analiza și evaluarea riscului de poluare a apei în industria textilă” [6].



Bibliografie

1. **DUȚU, M.:** Dreptul internațional al mediului, Ed. Economică, București, 1996.
2. **FILIP, F.G.:** Sisteme suport pentru decizii, Editura Tehnică, București, 2004.
3. **INMON, W.:** What is a Data Warehouse?, Prism Solutions, Inc., 1995.
4. **MINISTERUL APELOR ȘI PROTECȚIEI MEDIULUI:** Programul "România Curată", 2002.
5. **POWER, D.J.:** Decision Support Systems: Concepts and Resources for Managers. Quorum Books, Westport, Conecticut, 2002.
6. **RĂDULESCU, C-Z.:** Dezvoltare durabilă prin analiza și evaluarea riscului de poluare a apei în industria textilă, Raport de cercetare la proiect RELANSIN 2126, ICI, București, 2004.
7. **RĂDULESCU, C-Z.:** Managementul riscului prin analiza multidimensională a datelor. În: Revista Română de Informatică și Automatică, ICI, București, vol. 15, nr. 3, 2005 (în curs de apariție).
8. **TEICHERT, V., H. KARCHER:** Local Indicators of Sustainable Development. În: Addy, Tony: Theology and the Future of Work. Manchester: William Temple Foundation, 1998.
9. **WCDE, (World Commission on Environment and Development):** Our Common Future, Earthscan Publications, London, 1987.
10. **WIENER, J.L.:** What is Data Warehousing and what is Stanford Doing About it?, An Overview Talk Given in the Stanford DB Seminar series, Fall, 1997.