

**Academia Română
Secția Știința și Tehnologia Informației
Institutul de Cercetări pentru Inteligență Artificială**

Referat III

Sistem pentru asistarea deciziilor bazat pe descoperirea cunoștințelor din date: rezultate experimentale

Coordonator științific:

Acad. Florin FILIP

Doctorand:

Cornel LEPĂDATU

**București
2013**

CUPRINS

1. INTRODUCERE	5
2. SOLUȚII INFORMATICE PENTRU PROIECTAREA CONCEPTUALĂ A DEPOZITELOR DE DATE	9
2.1 Modelarea multidimensională a datelor	9
2.2 Proiectarea conceptuală a depozitelor de date	10
2.3 Metode orientate către date	12
2.4 Metode orientate către cerințe	12
2.5 Metode hibride	14
2.5.1 Metode hibride pure	14
2.5.2 Metode hibride paralele	15
2.5.3 Metode hibride secvențiale	15
3. STUDIU DE CAZ	23
3.1 Sistemul suport pentru decizii al unei biblioteci	23
3.1.1 Obiectivele sistemului	23
3.1.2 Direcții de îmbunătățire a activităților	23
3.1.3 Arhitectura sistemului	24
3.1.4 Resurse necesare	25
3.1.5 Avantajele sistemului	25
3.2 Analiza cerințelor informaționale	26
3.2.1 Cerințele biblioteconomice	26
3.2.1.1 Cerințe instituționale	26
3.2.1.2 Procese biblioteconomice	27
3.2.1.3 Măsurarea activităților	27
3.2.1.4 Indicatori operaționali	28
3.2.1.5 Indicatori de performanță	31

3.2.2 Cerințele bibliografice	39
3.2.2.1 Cerințe funcționale privind datele bibliografice	39
3.2.2.2 Definirea entităților și relațiilor	40
3.2.2.3 Descrierea entităților	43
3.2.2.4 Descrierea relațiilor	46
3.2.3 Cerințele bibliometrice	49
3.2.3.1 Indicatori bibliometrici	49
3.2.3.2 Indicatori bibliometrici de productivitate	51
3.2.3.3 Indicatori bibliometrici de performanță	53
3.2.3.4 Comentarii	61
3.3 Reconcilierea cu sursele de date	61
3.3.1 Sursele de date	61
3.3.2 Surogat bibliografic <i>documente</i>	65
3.3.3 Surogat bibliografic <i>publicații</i>	66
3.4 Depozitarea datelor	66
3.4.1 Identificare fapte	66
3.4.2 Definire dimensiuni	66
3.4.3 Definire ierarhii dimensionale	67
3.4.4 Definire măsuri	68
3.4.5 Setul de interogări preliminare	68
3.4.6 Cuburi de date	69
3.4.7 Schema conceptuală a depozitului de date	70
3.5 Descoperirea cunoștințelor din date	71
3.5.1 Nucleu de autoritate	72
3.5.2 Grupare bazată pe densitate	72
3.5.3 Programul CBD	75
2. CONCLUZII	83
BIBLIOGRAFIE	85
LISTA FIGURILOR	87
LISTA TABELELOR	87

1. INTRODUCERE

Conceptul de sistem suport pentru decizii desemnează 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 care interacționează cu celelalte părți ale sistemului informatic global al unei organizații. Menirea unui sistem suport pentru decizii este de a atenua efectul limitelor și restricțiilor decidentului intelectual într-un număr semnificativ de activități pentru rezolvarea unei palete largi de probleme decizionale nebanale pe baza implementării computerizate a unora dintre funcțiile de suport ale deciziilor care ar fi fost realizate altfel de către o echipă decizională ierarhică.

Dezvoltarea ideilor privind sistemul uman suport pentru decizii, privind modelul bazat pe cunoaștere al activităților decizionale, privind funcțiunile unui procesor pentru probleme decizionale sau privind extinderile posibile ale sistemelor de gestiune a bazelor de date pentru a integra cunoștințe descriptive (date) și cunoștințe procedurale (modele), au permis conturarea unui cadru conceptual generic sau arhitectură generică care să acopere majoritatea soluțiilor arhitecturale, identificabile în sisteme suport pentru decizii specifice, indiferent de domeniul de aplicație, de abordarea constructivă și de tehnologia informatică folosită.

Prin prisma arhitecturii generice, orice sistem suport pentru decizii se compune din patru componente esențiale: un sistem de limbaj, format din mesaje pe care sistemul le poate accepta; un sistem de prezentare, format din mesaje pe care sistemul le poate emite; un sistem al elementelor de cunoaștere, constând din cunoștințe deținute de sistem și, în fine, un sistem de tratare a problemei, constând din module software prin care elementele de cunoaștere sunt prelucrate ca urmare a interpretării mesajelor de intrare.

Arhitectura generică permite evidențierea diferențierilor dintre categoriile distincte de sisteme suport pentru decizii, arhitecturile personalizate păstrând caracteristicile sugerate de modelul conceptual generic dar fiind orientate către o anumită tehnologie (sau tehnologii) de reprezentare și prelucrare de cunoștințe.

În funcție de tehnologia dominantă, sistemele suport pentru decizii pot fi: orientate către texte, orientate către hipertext, orientate către baze de date, orientate către foi electronice de calcul, orientate către reguli sau orientate către rezolvatoare (solver). Dacă factorul decizional are nevoie de capacitățile de prelucrare oferite de mai multe tehnologii de management al cunoștințelor există două opțiuni de bază: utilizarea mai multor sisteme suport pentru decizii, fiecare orientat către o anumită tehnologie sau utilizarea unui singur sistem suport pentru decizii, dar care integrează mai multe tehnologii.

Un caz special de integrare, deosebit de important prin implicațiile sale, îl reprezintă combinația dintre o tehnologie de management a rezolvatoarelor flexibile și o tehnologie de management a bazelor de date. Cu toate că această arhitectură acoperă doar o parte din posibilitățile identificate de arhitectura generică ea este adesea citată, în cărți și articole de specialitate, ca fiind arhitectura sistemelor suport pentru decizii. Foarte utilizată în prezent este o variantă a acestei combinații respectiv integrarea depozitării datelor cu rezolvatoare analitice (prelucrare analitică on-line) și rezolvatoare data mining (mineritul datelor și descoperirea cunoștințelor).

Procesul de construire al unui sistem suport pentru decizii specific de aplicație se compune din o serie de activități care încep cu generarea ideii de introducere a sistemului în

organizație și se termină cu obținerea unei versiuni relativ stabile, utilizabile în mod curent, a sistemului. Etapele corespund ciclului de viață al oricărui sistem informatic, evoluția și perfecționarea sistemului continuând și după începerea folosirii acestuia în mod curent.

Principalele tehnologii informatice menite să susțină activitățile decizionale care presupun analiza datelor sunt: depozitarea datelor, prelucrarea analitică on-line, mineritul datelor și descoperirea cunoștințelor.

Conceptul de depozit de date desemnează o colecție de date orientate pe subiecte, integrate, istorice și nevolatile destinată sprijinirii procesului de luare a deciziilor manageriale.

Depozitarea datelor este un proces care constă, mai întâi, dintr-o fază de construire (prin integrare, curățire și consolidare a unor colecții de date) și apoi, dintr-o fază de utilizare (prin tehnologii integrate de asistare a deciziilor) a unui depozit de date, oferind factorilor decizionali ai unei organizații, arhitecturi și instrumente pentru a organiza sistematic, a înțelege și, mai ales, pentru a utiliza datele în luarea deciziilor.

Prelucrarea analitică on-line desemnează o categorie de instrumente software care permit analiștilor și managerilor să înțeleagă esența datelor printr-un acces rapid, consistent și interactiv la o mare varietate de vederi posibile ale informațiilor, care au fost obținute prin transformarea datelor primare, astfel încât să reflecte dimensiunile reale ale organizației așa cum o percepe și o înțelege utilizatorul. Esențial este faptul că este nevoie de a reprezenta în mod explicit caracteristici importante ale informațiilor, care nu mai sunt legate de reprezentarea abstractă a conceptelor lumii reale ci, mai degrabă, de obiectivul factorilor de decizie și anume susținerea proceselor de analiză a datelor orientate către luarea deciziilor.

Există cel puțin două noțiuni specifice pe care orice model conceptual pentru baze de date multidimensionale trebuie să le includă într-o anumită formă și anume faptul și dimensiunea. Faptul este o entitate, a unei aplicații, care face obiectul unei analize orientată către decizie, reprezentabilă prin intermediul cubului de date, iar dimensiunea corespunde perspectivei din care faptele pot fi analizate în mod concludent. Aspectele specifice și măsurabile ale unui fapt, relevante pentru analiză, sunt numite măsuri.

Eficiența modelării în depozitarea datelor depinde strict de capacitatea de a descrie datele factuale în funcție de dimensiunile adecvate, adică în funcție de perspectivele din care datele pot fi analizate. Pentru a putea susține mai bine analiza datelor este util ca pentru fiecare dimensiune să se organizeze o ierarhie de niveluri obținabilă prin gruparea elementelor dimensiunii în funcție de nevoile analizei. Un nivel are asociate, de obicei, attribute descriptive (nume și descrieri). O dimensiune are deci trei componente principale: o mulțime de niveluri, o mulțime de descrieri de niveluri și o ierarhie între niveluri.

O colecție de măsuri ale aceluiași fapt este reprezentată prin metafora „cub de date”, având câte o dimensiune „fizică” pentru fiecare dimensiune „conceptuală” a măsurării: o coordonată a cubului de date specifică o combinație de membri ai nivelurilor mai multor dimensiuni iar celula corespunzătoare conține măsura (măsurile) asociată (asociate) unei astfel de combinații. În cazul în care membrii unui nivel pot fi agregați în membri ai unui alt nivel se spune că primul nivel se agregă la cel de al doilea nivel.

Proiectarea conceptuală a unui depozit de date este pasul cel mai important în reprezentarea corectă a unui domeniu de interes, fiind elementul esențial asupra căruia atât factorii de decizie cât și informaticienii sunt de acord. Este foarte important ca proiectanții unui depozit de date să urmeze o metodologie de proiectare conceptuală, consolidată și robustă dat fiind că dezvoltarea unui depozit de date este un proces foarte scump chiar și astăzi când există multe instrumente software oferind soluții prefabricate care acoperă toate etapele din ciclul de viață al depozitului de date.

Proiectarea conceptuală a unui depozit de date se bazează pe metafora „cubului de date” care poate fi obținută prin mai multe categorii de metode: orientate către date, metode orientate către cerințe și metode mixte sau hibride.

În abordarea orientată către date schemele multidimensionale se definesc printr-un proces de reinginerie a surselor de date, reducând însă participarea utilizatorilor finali și riscând astfel un eșec posibil față de așteptările acestora. Abordarea orientată către date poate asigura o trasabilitate perfectă și consistență cu sursele de date - de fapt garantează prezența datelor ce vor fi utilizate în procesarea analitică - dar nu poate exclude lipsuri în raport cu cerințele utilizatorilor.

În abordarea orientată către cerințe schemele multidimensionale se definesc pe baza obiectivelor instituționale care rezultă din cerințele factorilor de interes. Sursele de date sunt luate în considerare ulterior, atunci când se proiectează extragerea, transformarea și încărcarea datelor și când conceptele multidimensionale (fapte, dimensiuni și măsuri) trebuie să fie mapate pe sursele de date pentru a programa planul de alimentare al depozitului de date. În acest moment însă se poate întâmpla ca proiectantul să descopere că datele necesare nu sunt disponibile în sursele de date. Abordarea orientată către cerințe permite proiectanților să obțină o schemă conceptuală foarte aproape de cerințele utilizatorilor dar care poate să nu fie susținută de disponibilitatea efectivă a datelor.

Aceste două abordări de proiectare conceptuală a depozitelor de date, orientată către date și, respectiv, către cerințe, sunt în antiteză una cu cealaltă proiectanții fiind obligați să aleagă una dintre ele. Pentru remedia neajunsurile fiecăreia dintre cele două abordări și pentru a valorifica avantajele fiecăreia dintre ele s-a recurs la definirea unor metode hibride bazate pe diferite modele multidimensionale.

Eforturile de a defini o strategie unitară care să integreze beneficiile celor două abordări au condus la abordarea hibridă, metodele elaborate putând fi grupate astfel: metode hibride pure, metode care efectuează procesul de proiectare luând în considerare simultan atât sursele de date cât și obiectivele instituționale și metode hibride integrate, metode care combină și integrează o etapă de abordare orientată către date cu o etapă de abordare orientată către cerințe. La rândul lor, metodele hibride pot fi grupate în: metode hibride paralele, în care cele două etape sunt executate independent și, la final, schemele multidimensionale provenind din fiecare etapă sunt comparate și integrate și metode hibride secvențiale, în care cele două etape sunt executate într-o ordine prefixată, ieșirea primei etape fiind utilizată ca intrare în a doua etapă.

Până în prezent rezultatele cele mai promițătoare au fost obținute prin metodele hibride secvențiale. Pe scurt, etapele generale ale unei astfel de metode sunt: analiza cerințelor informaționale, modelarea multidimensională, reconcilierea cu sursele de date, generarea arborilor de atribute și modelarea avansată a datelor.

Mineritul datelor și descoperirea cunoștințelor este un ansamblu de metode și algoritmi destinat explorării și analizei unor (adesea) mari volume de date în vederea deducerii, din aceste date, a unor reguli, a unor asocieri, a unor tendințe necunoscute (nefixate a priori), a unor structuri specifice care să restituie în mod concis esența informației utile pentru asistarea deciziilor.

Procesul de realizare al unui sistem suport pentru decizii de bibliotecă, nou și captivant, creează multe provocări dar promite mari îmbunătățiri în modul de desfășurare a activităților, în modul de înțelegere a ceea ce se face în prezent și a ceea ce se preconizează pentru viitor.

2. SOLUȚII INFORMATICE PENTRU PROIECTAREA CONCEPTUALĂ A DEPOZITELOR DE DATE

2.1 Modelarea multidimensională a datelor

Există în prezent o multitudine de soluții de modelare conceptuală a datelor, cu caracteristici și putere expresivă diferite în funcție, în principal, de domeniile de aplicație pentru care au fost elaborate.

În contextul depozitării datelor (data warehousing) esențial este faptul că, în proiectarea depozitului de date, este nevoie de a reprezenta în mod explicit caracteristici importante ale informațiilor, dar care nu mai sunt legate strict de reprezentarea abstractă a conceptelor lumii reale ci, mai degrabă, de obiectivul final al depozitului de date și anume susținerea proceselor de analiză a datelor orientate către luarea deciziilor [7, 13, 30].

Există cel puțin două noțiuni specifice pe care orice model conceptual pentru depozitarea datelor trebuie să le includă într-o anumită formă și anume *faptul* și *dimensiunea*. *Faptul* este o entitate a unei aplicații care face obiectul unei analize orientată către decizie. *Dimensiunea* corespunde perspectivei din care faptele pot fi analizate în mod concludent. Aspectele specifice și măsurabile ale unui fapt, relevante pentru analiză, sunt numite *măsuri*.

Eficiența modelării în data warehousing depinde strict de capacitatea de a descrie datele factuale în funcție de dimensiunile adecvate, adică în funcție de perspectivele din care datele pot fi analizate. Pentru a putea susține mai bine analiza datelor este util ca pentru fiecare dimensiune să se organizeze o *ierarhie de niveluri* obținabilă prin gruparea elementelor/membrilor dimensiunii în funcție de nevoile analizei. Un nivel are asociate, de obicei, *attribute descriptive* (*nume* și *descreri*). O dimensiune are deci trei componente principale: o mulțime de niveluri, o mulțime de descrieri de niveluri și o ierarhie între niveluri.

O colecție de măsuri ale aceluiași fapt este reprezentată prin metafora „cub de date”, având câte o dimensiune „fizică” pentru fiecare dimensiune „conceptuală” a măsurării: o *coordonată* a cubului de date specifică o combinație de membri ai nivelurilor mai multor dimensiuni iar celula corespunzătoare conține măsura (măsurile) asociată (asociate) unei astfel de combinații. În cazul în care membrii unui nivel λ pot fi agregați în membri ai unui alt nivel λ' se spune că (nivelul) λ *se agregă la* (nivelul) λ' sau λ *rolls-up to* λ' .

Fie dată o mulțime numărabilă de *attribute descriptive* și fie T o mulțime finită de *tipuri de valori de bază*, cum ar fi întreg, zecimal, dată calendaristică sau text. Fiecare tip de valori de bază $\tau \in T$ este asociat cu un domeniu $\text{dom}(\tau)$ de *valori de bază* ale acestui tip. Fiecare element \mathbf{a} al unei mulțimi obiecte \mathcal{O} din viața reală, $\mathbf{a} \in \mathcal{O}$, este identificat în mod unic prin utilizarea unor valori, $\text{id}(\mathbf{a})$, distincte de valorile de bază, numite *identificatori*.

Fie \mathbf{D} o mulțime finită de dimensiuni $\mathbf{D} = \{D\}$.

Definiția 1. O *schemă dimensională*, $\mathbf{D} = (\Lambda, \Delta, \rho)$, este formată din:

- ↳ o mulțime finită Λ de *nume de niveluri*, $\lambda \in \Lambda$;
- ↳ o mulțime finită Δ_λ de *descrieri de niveluri* pentru fiecare nivel $\lambda \in \Lambda$, $\cup_{\lambda \in \Lambda} \Delta_\lambda = \Delta$, fiecare descriere de nivel $\partial \in \Delta_\lambda \subseteq \Delta$ este asociată cu un tip de bază $\tau(\partial) \in T$;
- ↳ o relație $\rho \subseteq \Lambda \times \Lambda$ de ordine parțială între nivelurile $\lambda \in \Lambda$; dacă $(\lambda_1, \lambda_2) \in \rho$ se spune că λ_1 *rolls-up to* λ_2 .

Definiția 2. O instanțiere dimensională pentru $\mathbf{D} = (\Lambda, \Delta, \rho)$, $\mathbf{d} = (\{\mathcal{O}_\lambda\}_{\lambda \in \Lambda}; \{\mathcal{M}_\lambda\}_{\lambda \in \Lambda}; \Phi_\Delta; \Psi_\rho)$, este formată din:

- ↪ o mulțime finită \mathcal{O}_λ de obiecte (din lumea reală) pentru fiecare nivel $\lambda \in \Lambda$, fiecare obiect $\mathbf{a} \in \mathcal{O}_\lambda$ are asociat un identificator unic $\mathit{id}(\mathbf{a}) \equiv \mu$ numit *membru* al lui λ , $\mu \in \mathcal{M}_\lambda$;
- ↪ o funcție $\varphi_\partial : \mathcal{M}_\ell \rightarrow \mathit{dom}(\tau(\partial))$, de la membrii fiecărui nivel $\lambda \in \Lambda$ la domeniul tipului de bază $\tau(\partial) \in T$ asociat cu fiecare descriere de nivel $\partial \in \Delta_\lambda \subseteq \Delta$, $\varphi_\partial \in \Phi_\Delta$;
- ↪ o funcție *rolls-up* $\lambda_1 \rightarrow \lambda_2 \equiv \psi_{\lambda_1, \lambda_2} : \mathcal{M}_{\lambda_1} \rightarrow \mathcal{M}_{\lambda_2}$, de la membrii nivelului λ_1 la membrii nivelului λ_2 pentru fiecare pereche de niveluri $(\lambda_1, \lambda_2) \in \rho$; dacă $\mu_2 = \mathit{rolls-up}_{\lambda_1 \rightarrow \lambda_2}(\mu_1)$ spunem că μ_1 *rolls-up* μ_2 , $\psi_{\lambda_1, \lambda_2} \in \Psi_\rho$.

Definiția 3. O schemă multidimensională, $(\mathcal{D}, \mathcal{F})$, este formată din:

- ↪ o mulțime finită $\mathcal{D} = \{ \mathbf{D} = (\Lambda, \Delta, \rho) \}$ de *scheme dimensionale*;
- ↪ o mulțime finită $\mathcal{F} = \{ \mathcal{f} [A_1 : \lambda_1, \dots, A_n : \lambda_n] \rightarrow [M_1 : m_1, \dots, M_k : m_k] \}$ de *scheme cub de date*, unde:
 - ↳ \mathcal{f} este un nume;
 - ↳ fiecare A_i ($i = 1 \div n$) este un nume distinct numit *atribut* al lui \mathcal{f} ;
 - ↳ fiecare λ_i este un nivel al lui \mathcal{D} ;
 - ↳ fiecare M_j ($j = 1 \div k$) este un nume distinct numit *măsură* a lui \mathcal{f} ;
 - ↳ fiecare m_j este fie un tip de bază ($\tau \in \mathcal{T}$), fie un nivel al lui \mathcal{D} ($\lambda \in \Lambda$).

Fie: $(\mathcal{D}, \mathcal{F})$ o schemă multidimensională, fie $\mathbf{d} = (\{ \mathcal{O}_\lambda \}_{\lambda \in \Lambda}; \{ \mathcal{M}_\lambda \}_{\lambda \in \Lambda}; \Phi_\Delta; \Psi_\rho)$ o instanțiere a lui $\mathbf{D} \in \mathcal{D}$, $\mathbf{D} = (\Lambda, \Delta, \rho)$ și fie $\mathcal{f} [A_1 : \lambda_1, \dots, A_n : \lambda_n] \rightarrow [M_1 : m_1, \dots, M_k : m_k]$ o schemă cub de date din \mathcal{F} .

Definiția 4. O *coordonată (conceptuală)* c a lui \mathcal{f} în (instanțierea) \mathbf{d} este un tuplu de atribute ale lui \mathcal{f} , adică o funcție care asociază fiecărui atribut A_i un membru al nivelului λ_i care apare în \mathbf{d} .

Definiția 5. Un *fapt* al lui \mathcal{f} în (instanțierea) \mathbf{d} este un tuplu de măsuri ale lui \mathcal{f} , adică o funcție care asociază fiecărui nume de măsură M_j fie cu o valoare (dacă m_j este un tip de bază, $\tau \in T$) fie cu un membru din \mathbf{d} (dacă m_j este un nivel, $\lambda \in \Lambda$).

Definiția 6. O instanțiere a unei (DB-)scheme multidimensionale $(\mathcal{D}, \mathcal{F})$ este formată din:

- ↪ o instanțiere dimensională \mathbf{d} pentru fiecare schemă dimensională $\mathbf{D} \in \mathcal{D}$, ($\mathbf{d} \in \mathbf{d}$);
- ↪ o funcție parțială numită *cub de date* care asociază coordonate ale lui \mathcal{f} în \mathbf{d} cu fapte ale lui \mathcal{f} în \mathbf{d} pentru fiecare schemă cub de date $\mathcal{f} \in \mathcal{F}$.

2.2 Proiectarea conceptuală a depozitelor de date

Proiectarea conceptuală a unui depozit de date este pasul cel mai important în reprezentarea corectă a unui domeniu de interes, în această privință atât factorii de decizie cât și informaticienii sunt total de acord. Prin urmare este foarte important pentru proiectanții de depozite de date să urmeze o metodologie de proiectare conceptuală, consolidată și robustă

dat fiind că dezvoltarea unui depozit de date este un proces foarte scump chiar și astăzi când există multe instrumente software acoperind toate etapele din ciclul de viață al depozitului de date și oferind soluții prefabricate [37].

Proiectarea conceptuală a unui depozit de date se bazează pe metafora „cubului de date” care poate fi obținută prin mai multe categorii de metode:

- ↳ metode orientate către date (*data-driven* sau *supply-driven*);
- ↳ metode orientate către cerințe (*requirement-driven*, *demand-driven* sau *goal-oriented*);
- ↳ metode mixte sau hibride (*hybrid*).

În abordarea orientată către date schemele multidimensionale se definesc printr-un proces de reinginerie a surselor de date dar reducând participarea utilizatorilor finali și riscând astfel un eșec posibil față de așteptările acestora. Abordarea orientată către date poate asigura o trasabilitate perfectă și consistență cu sursele de date, de fapt garantează prezența datelor ce vor fi utilizate în procesarea analitică, dar nu poate exclude lipsuri în raport cu cerințele utilizatorilor.

Într-o abordare orientată către cerințe schemele multidimensionale se definesc pe baza obiectivelor instituționale care rezultă din cerințele factorilor de interes. Sursele de date sunt luate în considerare ulterior, atunci când se proiectează extragerea, transformarea și încărcarea datelor (*ETL - Extraction, Transformation, Loading*) și când conceptele multidimensionale (fapte, dimensiuni și măsuri) trebuie să fie mapate pe sursele de date pentru a programa planul de alimentare al depozitului de date. În acest moment, se poate întâmpla ca proiectantul să descopere că datele necesare nu sunt disponibile în sursele de date.

Abordarea orientată către cerințe permite proiectanților să obțină o schemă conceptuală foarte aproape de cerințele utilizatorilor dar care poate să nu fie susținută de disponibilitatea efectivă a datelor.

Aceste două abordări de proiectare conceptuală a depozitelor de date, orientată către date și orientată către cerințe, sunt în antiteză una cu cealaltă și proiectanții sunt obligați să aleagă una dintre ele. Pentru remedia neajunsurilor fiecăreia dintre cele două abordări și pentru a valorifica avantajele fiecăreia dintre ele s-a conturat ideea definirii unor metode hibride bazate pe diferite modele multidimensionale.

Eforturile de a defini o strategie unitară care să integreze beneficiile celor două abordări au condus la abordarea hibridă metodele definite putând fi grupate în:

- ↳ metode hibride pure, metode care efectuează procesul de proiectare luând în considerare simultan atât sursele de date cât și obiectivele instituționale [33];
- ↳ metode hibride integrate (*integration-derived*), metode care combină și integrează o etapă de abordare orientată către date cu o etapă de abordare orientată către cerințe și care, la rândul lor, pot fi grupate în:
 - ↳ metode hibride paralele, în care cele două etape sunt executate independent și, la final, schemele multidimensionale provenind din fiecare etapă sunt comparate și integrate [2, 32];
 - ↳ metode hibride secvențiale, în care cele două etape sunt executate într-o ordine prefixată, ieșirea primei etape fiind utilizată ca intrare în a doua etapă [11, 27, 29, 34].

2.3 Metode orientate către date

În abordarea orientată către date, metodele de proiectare a schemelor multidimensionale încep cu interviuarea utilizatorilor finali, în scopul de a identifica fapte și dimensiuni de interes și de a defini un set preliminar de interogări (*preliminary workload*). Cerințele utilizatorilor sunt reprezentate în limbaj natural, vag și informal. Cerințele identificate nu reprezintă constrângeri pentru proiectant ci doar un fel de sugestii posibil de folosit într-o fază ulterioară de *remodelare* a datelor, proiectarea conceptuală bazându-se în principal pe experiența și capacitatea proiectantului.

Metodele se concentrează în continuare pe analiza surselor de date, pe normalizarea și integrarea schemelor, eventual pe alte aspecte care trebuiesc rezolvate înainte de proiectarea conceptuală a depozitului de date. Proiectarea conceptuală a depozitului de date utilizează modelul *DFM (Dimensional Fact Model)* care definește cuburi și dimensiuni pornind de la schemele relaționale sau E/R ale surselor de date [12]. Metoda de modelare conceptuală este semiautomată bazându-se pe un algoritm care generează un arbore de atribute reprezentând o sursă de date integrată. Rădăcina arborelui este o relație sau o entitate care a fost aleasă în mod arbitrar de către proiectant drept fapt. Algoritmul generează câte un nod pentru fiecare atribut întâlnit în timp ce navighează recursiv relațiile și entitățile din sursele de date. Pentru proiectarea conceptuală a depozitului de date generarea arborelui de atribute reprezintă un prim rezultat.

Proiectantul are, în continuare, posibilitatea de a *remodela* arborele de atribute prin adăugări de noi noduri, prin eliminări de noduri inutile sau prin schimbări de noduri părinte.

Când remodelarea se consideră finalizată, arborele de atribute este interpretat ca un cub de date: rădăcina arborelui reprezintă faptul, un nod fiu de rădăcină (eventual prevăzut cu ierarhii) reprezintă o dimensiune iar un nod frunză legat direct de rădăcină reprezintă o măsură.

În general, etapele unei astfel de metode sunt următoarele:

- ↳ identificarea cerințelor utilizatorilor prin interviuri cu utilizatorii finali;
- ↳ analiza și integrarea surselor de date în scopul de a crea o schemă globală;
- ↳ din schema globală, crearea câte unui arbore de atribute pentru fiecare posibil fapt, pe baza experienței proiectantului și cerințelor utilizatorilor;
- ↳ remodelarea arborelui în scopul de a defini conceptele multidimensionale (dimensiuni și măsuri);
- ↳ crearea de scheme de fapte transformabile în scheme stea / fulg de nea.

Datorită dificultăților de înțelegere a nevoilor utilizatorilor, metodele orientate către date au o mare probabilitate de eșec în realizarea unui depozit de date care să sprijine în mod eficient luarea deciziilor. În plus, aceste metode pot conduce la un consum mare de timp, deoarece proiectantul nu își poate concentra atenția doar pe partea utilă a sursei de date ci este obligat să examineze întreaga sursă de date pentru a identifica elementele multidimensionale.

2.4 Metode orientate către cerințe

În abordarea orientată către date nu se oferă nici o metodă formală pentru a ține seama de cerințele utilizatorilor în timpul creării și remodelării arborelui de atribute. Datorită faptului că cerințele utilizatorilor sunt slab reprezentate și puțin utilizate, proiectantul poate avea mari

dificultăți privind identificarea alegerii "corecte" a rădăcinii arborelui de generat (în special în prezența unui număr mare de relații într-o sursă de date integrată), privind identificarea atributelor care nu sunt necesare, privind identificarea atributelor care trebuie adăugate sau privind determinarea numărului strict necesar de niveluri din ierarhia unei dimensiuni. Dificultățile de acest gen pot fi depășite în mod satisfăcător pe baza cerințelor utilizatorilor, ele reprezentând de fapt singurele constrângeri care pot ghida proiectantul în modelarea conceptuală.

Pentru a descrie activitățile și relațiile existente între actorii implicați în mediile de depozitare a datelor (factorii de decizie și depozitul de date în sine) metoda care s-a relevat eficientă și a și fost adoptată cu succes s-a bazat [26] pe *Cadrul-i** (*i* framework*) propus pentru a modela, în faza de definire a cerințelor, rolurile active ale actorilor unui sistem informatic. *Cadrul-i** permite proiectanților să realizeze analiza aprofundată a unui domeniu și să elaboreze un model formal al mediului decizional.

Acest model poate fi folosit ca punct de plecare pentru a crea o schemă multidimensională bazată exclusiv pe cerințele utilizatorilor:

- informațiile solicitate de factorii de decizie sunt convertite în elemente multidimensionale (fapte, măsuri și dimensiuni) reprezentate în conformitate cu formalismul *UML (Unified Modeling Language)* extins [22, 35] cu profile care să acopere fiecare aspect al ciclului de viață al depozitului de date;
- ↳ în *Cadrul-i** cerințele utilizatorilor, respectiv obiectivele instituționale, sunt reprezentate într-o ierarhie detaliată de obiective imbricate:
 - ↳ strategice, obiectivele de nivel înalt care urmează să fie atinse de către organizație;
 - ↳ decizionale, pentru a răspunde la modul în care obiectivele strategice pot fi îndeplinite;
 - ↳ informaționale, pentru a defini informațiile necesare pentru luarea deciziilor;
- ↳ din obiectivele informaționale sunt derivate cerințele informaționale adică un set preliminar de interogări de nivel înalt strict legate de conceptele multidimensionale;
- ↳ pentru a descrie relațiile dintre actorii principali ai organizației, în funcție de interesele fiecăruia, proiectantul elaborează [23] un model, *SDM (strategic dependency model)*, care evidențiază modul în care depozitul de date poate ajuta fiecare actor să își atingă obiectivele inclusiv când acestea prezintă interdependențe;
- ↳ pentru fiecare actor inclus în *SDM* proiectantul elaborează câte un model, *SRD (strategic rationale model)*, care descrie în mod detaliat obiectivele instituționale specifice și activitățile;
- ↳ modelul *SDM* și modelele *SRD* permit obținerea schemei multidimensionale finale a depozitului de date, în conformitate cu formalismul *UML*.

Proiectantului nu i se oferă însă posibilitatea de a *raфина* eventual astfel de scheme pentru a opera, atunci când este cazul, modificările necesare. Rafinarea ar putea, de exemplu, să impună includerea de măsuri calculate sau modificarea nivelurilor ierarhice ale dimensiunilor.

Abordarea orientată către cerințe permite proiectanților să obțină o schemă conceptuală foarte aproape de cerințele utilizatorilor care însă nu poate fi susținută de disponibilitatea efectivă a datelor.

2.5 Metode hibride

Schema multidimensională obținută printr-o abordare orientată către cerințe poate prezenta o structură dificil de populat deoarece corespondențele dintre elementele multidimensionale și elementele entităților din sursele de date pot fi greu de detectat. În cazurile cele mai nefavorabile elementele multidimensionale identificate nu au toate omologi în sursele de date, ceea ce înseamnă că factorul de decizie dorește informații care nu sunt susținute de date efective. Pentru a depăși astfel de impedimente și în scopul de a obține scheme multidimensionale în concordanță cu datele operaționale, o extensie a metodei sugerează *reconcilierea* schemelor multidimensionale *UML* cu sursele de date prin aplicarea formelor normale multidimensionale [20, 25].

Etapele de parcurs sunt:

- ↪ efectuarea unei analize profunde a domeniului pe baza cerințelor utilizatorilor în scopul de a identifica obiectivele instituționale;
- ↪ analiza obiectivelor instituționale, în scopul de a identifica cerințele informaționale;
- ↪ definirea schemelor multidimensionale, folosind cerințele informaționale;
- ↪ reconcilierea schemelor multidimensionale cu sursele de date;
- ↪ transformarea schemelor multidimensionale reconciliate în scheme stea/fulg de nea.

Schemele multidimensionale produse de această extensie nu numai că sunt orientate către cerințe dar asigură și trasabilitatea cu datele care urmează să fie utilizate în luarea deciziilor.

Această metodă îmbunătățită a fost ulterior prezentată ca și metodologie hibridă [27] pentru că într-adevăr, folosind aceste scheme multidimensionale, proiectantul gestionează o imagine simplificată și compactă a surselor de date.

Totuși, proiectantului nu i se oferă posibilitatea de a *rafina* eventual astfel de scheme pentru a opera, atunci când este cazul, modificările necesare.

2.5.1 Metode hibride pure

O metodă hibridă pură se bazează pe ipoteza că cerințele utilizatorilor ar putea fi în întregime acoperite prin definirea unui set preliminar de interogări (*preliminary workload*) care să conțină toate interogările analitice pe care utilizatorii finali ar intenționa să le execute pentru a obține din depozitul de date informațiile necesare [33].

În această ipoteză există elaborat un algoritm capabil să genereze în mod automat un graf, ale cărui noduri sunt tabele ale surselor de date și ale cărui arce sunt join-uri de tabele, pe care este posibil să se identifice dacă o anumită tabelă trebuie considerată fapt sau dimensiune și să fie etichetată ca atare. De exemplu, o tabelă a cărei cheie primară apare într-o clauză *group by* este etichetată ca nivel dimensional. Etichetele sunt atribuite prin examinarea rolurilor jucate de tabele și de attribute în setul de interogări preliminare. Etichetarea corectă a tuturor nodurilor generează în mod normal o schemă multidimensională valabilă.

Metoda ajută proiectanții, în primele etape ale procesului de proiectare, să identifice rapid și în mod automat faptele și dimensiunile din sursele de date și rezultatul obținut reprezintă un bun punct de pornire pentru *rafinări* suplimentare.

Totuși, interogările preliminare, care provin din obiective instituționale, nu pot avea în vedere nici sintactic, nici semantic sursele de date și nu există nici o mențiune despre cum să se procedeze în cazul unor eventuale incompatibilități sintactice sau semantice. De fapt,

algoritmul doar atribuie o etichetă pentru fiecare tabelă din sursele de date dar nu produce o nouă schemă.

2.5.2 Metode hibride paralele

În conformitate cu abordarea hibridă paralelă [2] cele două metode, metoda orientată către cerințe și metoda orientată către date, sunt aplicate independent una de alta și cu fiecare dintre ele se obține câte o schemă multidimensională distinctă:

- ↳ una derivată din paradigma *GQM* („goal, question, metric”) și
- ↳ cealaltă derivată dintr-o reinginerie a schemelor *E/R*.

Schema multidimensională finală se obține prin *maparea* celor două scheme multidimensionale conform cu o strategie de mapare care verifică dacă acestea evidențiază aceleași fapte și măsuri în raport cu setul de dimensiuni considerat.

2.5.3 Metode hibride secvențiale

Metodele hibride secvențiale pornesc cu analiza obiectivelor instituționale definite de către factorii de decizie [25, 27]. Folosind aceste obiective, primul rezultat obținut este o schemă de reprezentare a cerințelor informaționale. În continuare, prin identificarea faptelor și dimensiunilor din cerințele informaționale, se obține o schemă conceptuală inițială a depozitului de date. Pentru a lua în considerație și sursele de date schema conceptuală inițială este *reconciliată* cu schemele logice ale surselor de date prin aplicarea formelor normale multidimensionale.

Reconcilierea cu sursele de date a unei scheme multidimensionale, obținută prin o abordare orientată către cerințe, nu este suficientă, în sine, pentru a defini o metodă hibridă. Reconcilierea înseamnă doar a verifica dacă o schemă conceptuală inițială este în acord cu sursele de date, în timp ce factorii de decizie ar dori să obțină și anumite informații care nu pot fi efectiv disponibile din cauza lipsei lor din sursele de date. În abordările orientate către date, metodele respective permit proiectanților ca, printr-un proces de reinginerie a surselor de date (*remodelare*), să modifice manual dependențele funcționale dintr-o schemă multidimensională atât prin eliminarea de relații inutile cât și prin introducerea de relații utile (adăugare de dimensiuni la cuburile de date sau creare de ierarhii pentru de căile de agregare), de asemenea, proiectanții pot adăuga atribute derivate din datele existente (ca măsuri calculate).

Metoda hibridă secvențială, descrisă în [34], adoptă un model bazat pe grafuri ce permite proiectantului să construiască cu ușurință o schemă grafică, așa-numitul graf al depozitului de date, care poate fi mapat pe modele relaționale sau orientate obiect. Proiectantul poate mai întâi să obțină o schemă multidimensională, care se potrivește cel mai bine nevoilor utilizatorilor și apoi să verifice compatibilitatea acesteia cu schemele surselor de date. Pentru a verifica compatibilitatea pot fi folosite tehnici automate iar eventuale incompatibilități semantice care pot apare în timpul mapării pot fi rezolvate folosind o ontologie.

Metoda hibridă secvențială, descrisă în [29], se concentrază mai întâi asupra surselor de date și ulterior asupra cerințelor utilizatorilor și se bazează pe un algoritm iterativ care identifică entitățile reprezentând fapte pe baza atributelor lor numerice. În prima iterație a algoritmului este analizată entitatea cu cel mai mare număr de atribute numerice, următoarele iterații luând în considerație, în ordine descrescătoare, celelalte entități. Dacă entitatea considerată participă într-o relație $n:1$, atunci această relație permite identificarea entităților ce

reprezintă dimensiuni și prin examinarea recursivă a tipului cardinalității relațiilor sunt construite nivelurile ierarhice ale dimensiunilor. În fiecare iterație a algoritmului se obține câte o schemă multidimensională cu un fapt candidat și dimensiunile sale. Algoritmul se termină când au fost analizate toate entitățile din sursele de date. Proiectantul poate alege schema multidimensională care permite cel mai bun răspuns la setul preliminar de interogări utilizând un tabel având atâtea rânduri câte interogări preliminare au fost definite și atâtea coloane câte scheme multidimensionale au fost create automat. Celula (i, j) din tabel conține simbolul "x" dacă interogarea i poate fi executată pe schema multidimensională j, simbolul "p" dacă schema multidimensională j răspunde parțial la interogarea i și " " (simbolul blank) în rest. În cazul în care schema multidimensională selectată nu răspunde complet toate interogările (adică rândul corespunzător nu conține peste tot "x") proiectantul poate rafina manual schema pe baza propriilor cunoștințe cu privire la cerințele utilizatorilor.

Metoda hibridă secvențială, descrisă în [11], începe cu analiza cerințelor pe baza metodologiei Tropos (*Tropos: an agent-oriented software development methodology*) și are ca scop de a modela domeniul de interes în funcție de două perspective diferite: modelarea organizațională, care descrie obiectivele factorilor de interes și modelarea decizională, care descrie obiectivele factorilor de decizie. Ulterior schemele multidimensionale derivate din analiza cerințelor se potrivesc cu cele derivate din sursele de date, în trei etape:

- ↳ în prima etapă conceptele multidimensionale identificate sunt mapate pe entitățile din sursele de date, rolul cel mai important este jucat de către schemele multidimensionale derivate din modelarea decizională cele provenite din modelarea organizațională fiind mai puțin utile;
- ↳ în a doua etapă, în scopul de a construi în mod automat ierarhiile pentru fiecare fapt mapat cu succes, sunt navigate recursiv relațiile $m : I$ și sunt generate scheme de fapte, în general mai simple și mai reduse decât dacă ar fi fost generate luând în considerare în mod direct sursele de date;
- ↳ în a treia etapă, proiectantul obține schema multidimensională finală prin rafinarea manuală a schemelor de fapte pentru a le adapta mai bine cerințelor utilizatorilor.

Evaluările actuale din literatura de specialitate sunt în acord cu privire la importanța covârșitoare a efortului (inclusiv financiar) pentru a identifica în mod corect nevoile utilizatorilor și cu privire la necesitatea crucială a reprezentării formalizate a cerințelor instituționale în cadrul procesului de proiectare conceptuală a depozitelor de date [37].

Proiectantul trebuie să ia în considerare întotdeauna atât schema multidimensională reprezentând sursele de date cât și schema multidimensională reprezentând cerințele instituționale și să asigure *aliniera* acestor scheme conceptuale în jurul acelorași elemente multidimensionale.

Pentru a obține alinierea, simpla *mapare* [25] nu este suficientă din moment ce trebuie făcute modificări ori de câte ori cerințele instituționale sunt inconsistente în raport cu sursele de date. Pentru a rezolva inconsistențele, se efectuează o *reconciliere* a schemelor multidimensionale cu sursele de date prin aplicarea formelor normale multidimensionale. De exemplu, pe baza primei condiții a primei forme normale multidimensionale, conform căreia fiecare dependență funcțională din schema multidimensională trebuie să corespundă cu o dependență funcțională din sursele de date, o anumită asociere trebuie eliminată sau, pe baza celei de a doua condiții a primei forme normale multidimensionale care prevede că fiecare dependență funcțională din sursele de date trebuie să fie reprezentată printr-o asociere în schema multidimensională, o anumită asociere trebuie adăugată. Reconcilierea este o abordare mai bună [20, 27]. Pe de o parte, schema multidimensională reconciliată este destul

de îndepărtată de schema multidimensională rezultată inițial din cerințele utilizatorilor pentru că reconcilierea a modificat schema originală și a eliminat anumite operații (*roll-up*, *drill-down*) dorite inițial de către factorii de decizie. Pe de altă parte, schema multidimensională reconciliată este foarte apropiată de sursele de date.

Când în urma alinierii se obține o schemă conceptuală consistentă proiectantul dispune de o vedere simplificată asupra surselor de date esențiale și poate apoi continua cu o *rafinare* manuală suplimentară pentru a asigura corectitudinea și completitudinea schemei conceptuale. Rafinarea are drept scop de a modifica schema multidimensională în conformitate cu o abordare orientată către date complete.

Dacă proiectantul dorește o schemă conceptuală care să se potrivească mai bine cu cerințele utilizatorilor, atunci schema multidimensională trebuie să fie *remodelată* manual (*advanced data modelling*) prin impunerea unor dependențe funcționale care nu sunt prezente în schema inițială dar cu condiția ca acestea să fie inerente domeniului de aplicație [13].

În scopul de a valorifica cele mai bune caracteristici ale fiecăreia dintre metode a fost dezvoltată o metodă hibridă secvențială [37] care acoperă atât analiza cerințelor cât și proiectarea conceptuală.

S-a dorit obținerea de scheme multidimensionale care, pe de o parte, să răspundă nevoilor utilizatorilor iar, pe de altă parte, să și permită proiectanților să execute activitatea de *remodelare* a datelor. Ideea principală a fost de a folosi ca intrări în faza de remodelare (sau modelare avansată a datelor) schemele multidimensionale *UML* reconciliate, obținute din *Cadrul-i** și modelarea multidimensională *UML*.

Metoda se bazează deci pe două modele multidimensionale:

- ↳ modelul *UML* pentru a reprezenta schemele multidimensionale orientate către cerințele depozitului de date și
- ↳ modelul *E-DFM* (*Extended Dimensional Fact Model*) pentru a reprezenta o viziune arborescentă asupra surselor de date.

Această viziune și anume arborele de attribute, permite proiectantului să opereze cu ușurință pe structura datelor esențiale din sursele de date. În conformitate cu acest model, modificarea dependențelor funcționale din sursele de date corespunde la operațiuni intuitive pe arbore, cum ar fi adăugarea și/sau eliminarea de noduri.

Modelarea multidimensională urmează metodologia hibridă convențională pentru a produce scheme *UML* reconciliate iar modelarea avansată a datelor (sau remodelarea) se bazează pe modelul *E-DFM* un model definit pentru a îmbunătăți anumite caracteristici ale modelului *DFM* și echipat cu un program logic capabil de a construi arbori de attribute din orice schemă relațională care conține relații binare.

Pentru a utiliza schemele multidimensionale *UML* reconciliate ca intrări pentru activitatea de modelare avansată a datelor, proiectantul se confruntă [1] cu o problemă de traducere de schemă conceptuală adică: fiind date două modele *M1* (*UML*) și *M2* (*E-DFM*) și o schemă multidimensională *S1* pe *M1* să se găsească o schemă multidimensională *S2* pe *M2* echivalentă cu *S1*.

Deși cele două modele sunt destul de diferite, atât în terminologie cât și în elementele grafice, ele permit reprezentări ale acelorași elemente multidimensionale în sensul că fiecare element din *UML* are omologul său în *E-DFM* și vice-versa.

Pentru a face traducerea schemei conceptuale sunt necesare:

- ↳ definirea unei mapări între elementele celor două modele multidimensionale;

↪ dezvoltarea un algoritm care generează arbori de atribute pornind de la scheme *UML*.

Fie x și respectiv y nodurile din arborele de atribute corespunzătoare atributelor X și respectiv Y . Pe un arbore de atribute generat sunt definite următoarele operații de bază:

- ↪ *create x*, crearea nodului x ;
- ↪ *delete x*, ștergerea nodului x ;
- ↪ *add $x \rightarrow y$* , adăugarea unui arc de la x la y ;
- ↪ *remove $x \rightarrow y$* , eliminarea arcului de la x la y .

Pot fi, de asemenea, definite operații complexe precum:

- ↪ *prune x*, ștergerea nodului x împreună cu toți copiii al căror părinte este;
- ↪ *graft x*, ștergerea nodului x și adăugarea copiilor săi la părinții nodului x ;
- ↪ *change parent of y from x to z* , care înseamnă că, având arcul $x \rightarrow y$ și nodul z ,:
 - ↳ *remove $x \rightarrow y$* ;
 - ↳ *add $z \rightarrow y$* .

În consecință, cele patru operații de bază definite pe arborele de atribute corespund respectiv la: crearea unui atribut x , ștergerea unui atribut x , adăugarea unei dependențe funcționale $x \rightarrow y$ și eliminarea unei dependențe funcționale $x \rightarrow y$. În plus, operațiunea de schimbare a părintelui este foarte utilă pentru a modifica nivelurile ierarhice ale dimensiunilor. Prin urmare, operațiile de bază permit efectuarea reingineriei schemelor multidimensionale folosind o abordare orientată către date completă.

În *UML*, fiecare element grafic este reprezentat cu un profil specific. Pentru a putea folosi schemele multidimensionale *UML* ca intrări pentru modelarea avansată a datelor este necesară definirea corespondențelor dintre elementele grafice ale modelelor multidimensionale *UML* și *E-DFM*. Fiecare element al *E-DFM* este întotdeauna afișat ca un nod, dar semantica acestuia este stabilită de poziția nodului în arbore. Astfel, rădăcina este *fapt*, nodurile frunză sunt *attribute descriptive* (cele legate direct la rădăcină sunt *măsuri*) iar nodurile care nu sunt frunze sunt *attribute dimensionale* (cele legate direct de rădăcină sunt *dimensiuni*, celelalte sunt *niveluri ierarhice* ale dimensiunilor).

Generarea arborelui de atribute (extins) din o schemă multidimensională *UML* este realizată de un algoritm, descris în continuare, care este capabil să genereze arbori de atribute conform cu modelul *E-DFM*. Algoritmul are nevoie de doar două primitive grafice:

- ↪ *root (x)*, funcția care creează un graf cu x ca nod rădăcină, și
- ↪ *add (y, z)*, funcția care creează pe y ca nod copil al nodului z .

Arborele este realizat ca un graf orientat, arcul este un drum (de obicei, reprezentat prin o săgeată normală) de la z la y . În cazul în care nodul y deja există, atunci este creat numai drumul de la z la y . În scopul de a distinge raportul de cardinalitate al relațiilor dintre noduri, se folosește o altă primitivă grafică, numită *add_dot (y, z)*, care creează o săgeată neorientată de la z la y dacă z *Rolls-upTo* y cu multiplicitatea n , adică, există o relație $m : n$ între z și y .

Pseudo-codul algoritmului este prezentat în Figura 1, unde:

- ↪ *base(d)* este funcția care extrage prima bază (adică cel mai scăzut nivel de agregare) din *Dimension class d*;
- ↪ *explore(z, y)* este funcția recursivă descrisă mai jos;
- ↪ *descriptor(b)* este funcția care extrage *Descriptor*-ii lui *Base class b*, și

↳ $cardinalityRolls-upTo(b, u)$ este funcția care returnează cardinalitatea asocierii *Rolls-upTo* dintre nodurile u și b .

Generarea arborelui se bazează pe funcția recursivă $explore(z, y)$ care navighează prin *Base classes* prin intermediul elementelor *Rolls-upTo* ale schemei UML. (*Base classe* z și *Base classe* y sunt astfel încât y are o relație *Rolls-upTo* cu z , adică este posibil să se ajungă de la *Base class* z la *Base class* y printr-o relație $1 : n$)

De fiecare dată când este întâlnită o *Base class*, este creat un nod, împreună cu copiii săi. Recursivitatea se termină când este întâlnită o *Base class* fără nici un element *Rolls-upTo*.

<i>Elemente UML</i>	<i>Notatii</i>	<i>Elemente E-DFM</i>
Fact class	- f <i>fapt</i>	Fact
Descriptor	- h <i>descriptor</i>	Dimensional attribute
Fact attribute	- m <i>măsură</i>	Measure
	- f_m <i>mulțimea de măsuri ale lui f</i>	
Dimension class	- d <i>dimensiune</i>	Dimension
	- f_d <i>mulțimea de dimensiuni ale lui f</i>	
Base class	- b <i>nivel dimensional de bază</i>	Dimensional level
Descriptor	- k <i>descriptor</i>	Dimensional attribute
Dimension attribute		Descriptive attribute
	- b_a <i>mulțimea de atribute ale lui b</i>	
	- u, v <i>nodes</i>	
Descriere algoritm (pseudocod)		
	$u = root(f)$	
	// adăugarea unei măsuri la rădăcină	
	for each m in f_m	
	$add(m, u)$	
	end for	
	for each d in f_d	
	$b = base(d)$	
	$explore(b, u)$	
	end for	
	function $explore(b, u)$	
	$k = descriptor(b)$	
	$h = descriptor(u)$	
	$n = cardinalityRolls-upTo(b, u)$	
	// adăugare atribut dimensional	
	if ($n = 1$)	
	$add(k, h)$	
	else	
	$add_dot(k, h)$	
	// adăugare atribut descriptiv	
	for each v in b_a	
	$add(v, k)$	
	end for	

	for each c such that b Rolls-upTo c	
	<i>explore</i> (c, b)	
	end for	
	end function	

Figura 1. Algoritm de generare a arborilor de atribute

În general, în metodele orientate către date, prima activitate este de a identifica în mod corect faptele din sursele de date [13]. Cu această metodă activitatea nu mai este necesară pentru că toate faptele au fost deja identificate și reprezentate folosind schemele multidimensionale *UML*. Totuși, pentru a putea efectua remodelarea schemelor multidimensionale *UML*, acestea trebuie redefinite în funcție de activitățile de bază ale unei metode orientate către date.

Procesul de remodelare se bazează nu numai pe experiența proiectantului ci și pe cerințele utilizatorilor. Astfel activitățile de modelare avansată a datelor pe arbori de atribute derivați din scheme multidimensionale *UML* reconciliate sunt:

- ↪ Definierea dimensiunilor: Pentru fiecare resursă, referitoare la un rol al actorului depozit de date, proiectantul alege dimensiunile unui fapt în funcție de diferite contexte de analiză. În cazul în care nu sunt definite ca atare în schemele multidimensionale *UML* aceste contexte de analiză ar trebui să devină copii ai unei rădăcini. Pot apare diferite situații. În situația în care contextul selectat corespunde la un nod în arbore, proiectantul efectuează o operație de schimbare de părinte pentru a-l face un copil al rădăcinii. În situația în care în arbore nu există nici un nod care să corespundă contextului respectiv, proiectantul nu poate adăuga această dimensiune la fapt din cauza lipsei efective a datelor din sursele de date și va trebui să caute alte contexte de analiză pentru a descrie faptul la un alt nivel de granularitate. Dacă nici această operațiune nu reușește atunci această dimensiune nu poate fi aleasă pentru acel fapt. Pe de altă parte copiii unei rădăcini care nu s-au dovedit necesari ca dimensiuni sunt eliminați din arbore.
- ↪ Definierea măsurilor: În cazul în care dimensiunile unui fapt se schimbă - în primul rând, această schimbare se întâmplă în timpul etapei de reconciliere - proiectantul trebuie să verifice din nou formele normale multidimensionale. Mai precis, trebuie să verifice dacă fiecare măsură derivată poate fi calculată pe baza datelor disponibile și dacă dimensiunile determină funcțional măsurile. Cu alte cuvinte, măsurile trebuie să fie corect reprezentate la nivelurile de agregare stabilite de setul de dimensiuni ale unui fapt.
- ↪ Definierea ierarhiilor: Pentru fiecare resursă referitoare la un rol al actorului depozit de date, toate contextele de analiză care nu sunt definite ca dimensiuni trebuie să formeze o ierarhie. În primul rând proiectantul alege drumul de agregare dorit după care fiecare context de analiză devine un nivel dimensional în ierarhie adică într-un sub-arbore având drept rădăcină dimensiunea respectivă. În cazul în care un nivel dimensional corespunde unui nod prezent în arbore proiectantul efectuează o operație de schimbare de părinte pentru a face nivelul dimensional n copil al nivelului dimensional $n - 1$. De exemplu, al doilea nivel dimensional devine copil al dimensiunii, al treilea nivel dimensional devine copil la al doilea și așa mai departe. Dacă nu există nici un nod în arbore care să corespundă unui nivel dimensional atunci contextul de analiză dorit nu poate fi furnizat pentru acel fapt. După ce toate contextele de analiză au fost luate în

considerare nodurile care nu corespund nici unui context de analiză sunt eliminate din arbore.

- ↳ Definierea atributelor descriptive: Proiectantul nu poate introduce atribute descriptive noi, pentru care datele necesare ar putea să nu fie disponibile și prin urmare el trebuie să se bazeze doar pe atributele capturate de către schemele *UML* și mapate cu succes pe sursele de date în pasul de reconciliere. Proiectantul poate însă elimina atributele descriptive inutile sau poate fuziona mai multe noduri într-un singur nod.
- ↳ Optimizarea bazată pe cunoaștere: Aceasta este o activitate opțională în care proiectantul poate modifica structura arborelui prin utilizarea experienței personale și a propriilor cunoștințe. Proiectantul poate îmbunătăți drumurile de agregare în ierarhii deoarece unele contexte de analiză necesare pot fi obținute, chiar dacă acestea nu sunt prezente în mod explicit în sursele de date. De exemplu, deoarece o dimensiune timp completă poate fi derivată dintr-o dată, proiectantul ar putea introduce semestrul ca nivel dimensional în ierarhia dimensiunii timp dacă acest lucru oferă un context de analiză interesant.
- ↳ Validarea: Presupunând că sarcinile descrise în modelele *SRM (Strategic Rationale Models)* ale factorilor de decizie corespund analizelor tipice care urmează să fie efectuate pe depozitul de date, proiectantul verifică dacă elementele multidimensionale din arborii de atribute oferă răspunsuri la aceste întrebări. Pentru fiecare fapt dat proiectantul trebuie să verifice dacă agregările cerute sunt operațiuni corecte și executabile. Pentru fiecare agregare proiectantul trebuie să se asigure că există un nivel dimensional corespunzător și pentru fiecare măsură, un atribut numeric corespunzător ca și copil al rădăcinii arborelui.

Pe scurt, etapele generale ale metodei sunt:

- ↳ Analiza cerințelor: constă în efectuarea, utilizând *Cadrul-i**, a unei analize profunde a domeniului pe baza cerințelor utilizatorilor, în scopul de a detecta obiectivele instituționale. Analiza obiectivelor instituționale permite apoi proiectantului să deducă cerințele informaționale.
- ↳ Modelarea multidimensională: constă în definirea manuală a schemelor multidimensionale utilizând cerințele informaționale.
- ↳ Reconcilierea: constă în reconcilierea schemelor multidimensionale prin maparea lor pe sursele de date și soluționarea inconsistențelor. Această activitate se poate face și automat [24] cu ajutorul instrumentului *QVT (Query-View-Transformation Facility)*.
- ↳ Generarea arborelui de atribute: constă în crearea în mod automat de arbori de atribute pentru fiecare schemă *UML* obținută în pasul anterior.
- ↳ Modelarea avansată a datelor: permite proiectantului remodelarea manuală a arborelui de atribute în scopul de a modifica dependențe funcționale. Acest lucru ar putea fi necesar pentru a alinia mai bine schema la cerințele utilizatorilor și pentru a îmbunătăți relațiile ierarhice dintre nivelurile dimensionale.

Activitatea de modelare avansată a datelor se face după modelarea multidimensională pe baza cerințelor utilizatorilor și din moment ce sunt utilizate schemele multidimensionale *UML* și nu scheme *E/R*, nivelul de abstractizare al reprezentării intrării în modelarea avansată a datelor nu a fost modificat.

Din cauza etapei de reconciliere, o schemă multidimensională *UML* reconciliată asigură disponibilitatea efectivă a datelor necesare. Chiar dacă arborele generat ar putea fi considerat

un surrogat al surselor de date el acoperă partea din sursele de date strict necesară pentru cerințele informaționale și prin urmare, proiectantul se va putea concentra în continuare numai pe relațiile relevante din sursele de date.

Metoda oferă proiectantului o schemă preliminară care reprezintă sursele de date, dar în care cerințele informaționale legate de obiectivele instituționale au fost complet acoperite. Utilizând ca intrare, pentru modelarea avansată a datelor, schema multidimensională *UML* reconciliată proiectantul poate să introducă măsuri derivate și poate să modifice dependențe funcționale pe baza unor considerente și a unei experiențe proprii. Pentru a exemplifica, dimensiunile care determină un fapt pot fi modificate și nivelurile de agregare într-o ierarhie, de asemenea. Așa cum s-a menționat anterior, acest lucru poate fi realizat nu numai în scop de optimizare (ca un exemplu, două atribute descriptive numele și prenumele se pot uni în unul singur pentru a reduce numărul de atribute), dar, de asemenea, pentru a asigura cea mai bună aderență a schemei la cerințele utilizatorului.

Alte avantaje ale acestei metode sunt după cum urmează.

- ↪ Schemele multidimensionale *UML* reprezintă o formalizare puternică a obiectivelor instituționale și oferă un limbaj comun atât pentru informaticieni cât și pentru factorii de decizie. Prin urmare, este mai greu să apară neînțelegeri și/sau omisiuni.
- ↪ Schemele multidimensionale *UML* reprezintă concepte multidimensionale bine definite, cum ar fi faptele și dimensiunile. Ca urmare, proiectantul nu mai este obligat să se ocupe cu identificarea directă a acestor concepte în sursele de date.
- ↪ Generarea schemelor multidimensionale *UML* reconciliate determină o selecție preliminară a atributelor, furnizând proiectantului numai pe cele strict necesare. Dacă ar lucra direct pe sursele de date proiectantul ar trebui să facă aceasă selecție pe foarte multe entități având, la rândul lor, multe atribute, cele mai multe dintre acestea nefiind utile.

3. STUDIU DE CAZ

3.1 Sistemul suport pentru decizii al unei biblioteci

Concepția și implementarea sistemelor informatice, deci și a sistemelor suport pentru decizii, sunt influențate [3] de către o serie de factori, printre care pot fi menționați: obiectivele urmărite, persoanele implicate, recomandările, normele și standardele utilizate, restricțiile impuse de către instituție, evoluția mediului, bugetul disponibil și termenele de finalizare.

3.1.1 Obiectivele sistemului

Provocările cu care se confruntă un sistem suport pentru decizii de bibliotecă [19] sunt:

- ↻ Îmbunătățirea satisfacției utilizatorilor (în sediul propriu și în afara lui);
- ↻ Elaborarea unei politici de achiziție orientată către cerere;
- ↻ Îmbunătățirea conservării colecțiilor;
- ↻ Optimizarea fluxurilor;
- ↻ Optimizarea alocării resurselor (umane și financiare);
- ↻ Diversificarea și creșterea veniturilor (culturale și comerciale);
- ↻ Comunicare mai bună cu partenerii (organisme guvernamentale, alte biblioteci).

Obiectivele sistemului suport pentru decizii de bibliotecă sunt:

- ↻ Furnizarea de indicatori care să permită evaluarea în timp a conformității cu obiectivele bibliotecii, prin:
 - ↳ evaluarea rezultatelor obținute, sesizarea tendințelor, alerte;
 - ↳ evaluarea indicatorilor operaționali;
 - ↳ evaluarea indicatorilor de performanță;
 - ↳ rapoarte de activitate.
- ↻ Furnizarea unor instrumente de analiză a tendințelor, de sesizare a situațiilor decizionale și de sugerare a unor acțiuni corespunzătoare (analize complexe, simulări, prognoze) în vederea luărilor de decizii.
- ↻ Integrarea datelor și compararea informațiilor din aplicații informatice existente.
- ↻ Simplificarea accesului la informație prin:
 - ↳ schimb transparent de informații;
 - ↳ diseminare accelerată a informațiilor.

3.1.2 Direcții de îmbunătățire a activităților

Utilizatorii sistemului pot solicita un spectru larg de expertize, de la căutări simple la statistici avansate. Pentru a putea adapta serviciile oferite de sistem la cerințele fiecărei categorii de utilizatori se au în vedere patru categorii principale de astfel de servicii:

- ↪ Căutări simple care furnizează rapoarte predefinite și indicatori de performanță;
- ↪ Interogări avansate și/sau interogări personalizate, la cerere;
- ↪ Analize avansate implicând navigare multidimensională și funcții puternice de analiză;
- ↪ Simulări și statistici avansate.

Domeniile posibile de îmbunătățire a activității bibliotecii sunt:

- ↪ Dezvoltarea colecțiilor;
- ↪ Accesibilitatea colecțiilor;
- ↪ Accesul la publicații;
- ↪ Utilizarea bibliotecii;
- ↪ Digitalizarea colecțiilor;
- ↪ Serviciile bibliografice, asistență, îndrumare;
- ↪ Potențialul de dezvoltare;
- ↪ Conservarea colecțiilor;
- ↪ Managementul.

3.1.3 Arhitectura sistemului

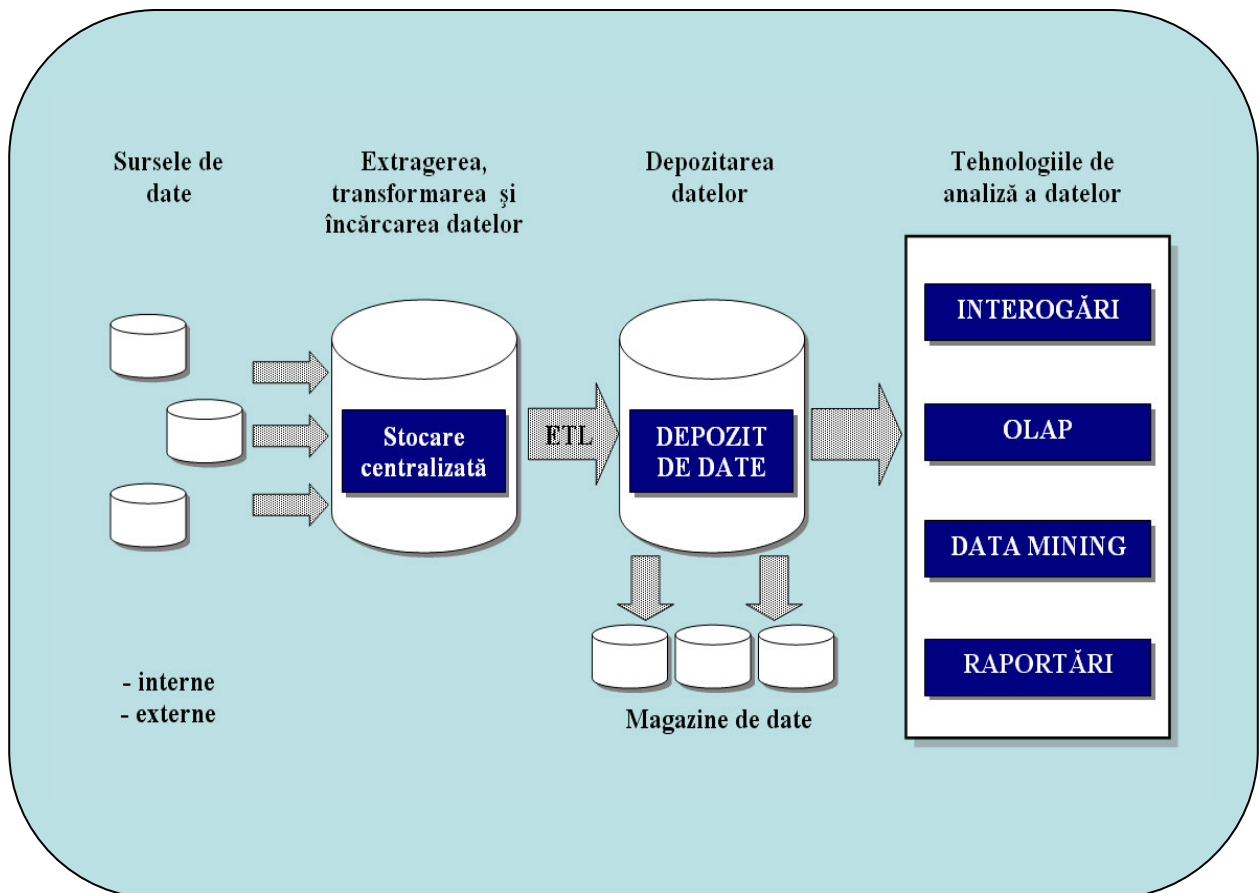


Figura 2. Arhitectura sistemului suport pentru decizii

Această arhitectură, combinație dintre o tehnologie de management a rezolvatoarelor flexibile și o tehnologie de management a bazelor de date, deși acoperă doar o parte din posibilitățile identificate de arhitectura generică [9, 10] ea este adesea citată, în cărți și articole de specialitate, ca fiind arhitectura sistemelor suport pentru decizii [5, 6]. În Figura 2 este prezentată o variantă a acestei combinații respectiv integrarea depozitării datelor cu rezolvatoare analitice (prelucrare analitică on-line) și rezolvatoare data mining (mineritul datelor și descoperirea cunoștințelor).

3.1.4 Resurse necesare

↪ Pentru resursele umane se impune următoarea structură:

- ↳ Un comitet de management;
- ↳ O echipă de proiect;
- ↳ Grupuri de lucru cu utilizatorii;
- ↳ O firmă de consultanță (pentru analiza cerințelor);
- ↳ Un subcontractant (pentru dezvoltare-implementare).

↪ Pentru resursele financiare trebuie avute în vedere:

- ↳ Instrumentele de fundamentare a deciziilor;
- ↳ Un instrument ETL pentru date ;
- ↳ Un instrument de raportare;
- ↳ Consultanță;
- ↳ Contract pentru dezvoltarea sistemului.

3.1.5 Avantajele sistemului

- ↪ Asigură informații de calitate și beneficiază de noi instrumente de management;
- ↪ Adaptabilitate, pe măsura utilizării sistemului cerințele pot evolua;
- ↪ Asigură satisfacerea cerințelor utilizatorilor și a cerințelor tehnice actuale și viitoare;
- ↪ Rezolvă o serie de faze tehnice critice privind furnizarea, modelarea și stocarea datelor;
- ↪ Impune personalului dezvoltarea în consecință a abilităților;
- ↪ Susține trecerea la o cultură orientată către performanță;
- ↪ Procesul de realizare a unui astfel de sistem, nou și captivant, creează multe provocări dar promite mari îmbunătățiri în modul de desfășurare a activităților, în modul de înțelegere a ceea ce se face în prezent și a ceea ce se preconizează pentru viitor.

3.2 Analiza cerințelor informaționale

3.2.1 Cerințele biblioteconomice

Știința care se ocupă cu studiul cărților sub aspectul publicării, difuzării și comercializării lor este bibliologia. Biblioteconomia, ramură a bibliologiei, se ocupă cu formarea, administrarea și organizarea bibliotecilor. Fondator al biblioteconomiei moderne poate fi considerat Gabriel Naudé (1600 - 1653), bibliotecarul cardinalului Mazarin, prin lucrarea sa *Advis pour dresser une bibliothèque* (1627). Recent, termenul de „biblioteconomie” este tot mai frecvent înlocuit prin „științele informării și bibliotecilor” (*Library and Information Science*).

Colecțiile unei biblioteci pot cuprinde următoarele categorii de documente: cărți, publicații seriale, manuscrise, microformate, documente cartografice, documente de muzică tiparite, documente audiovizuale, documente grafice, colecții electronice, documente fotografice sau multiplicare prin prelucrări fizico-chimice, documente arhivistice, alte categorii de documente indiferent de suportul lor material.

3.2.1.1 Cerințe instituționale

Principalele cerințe la care trebuie să răspundă o bibliotecă națională [21] se referă la:

- ↳ Colectarea și prezervarea publicațiilor de interes național general și de importanță deosebită:
 - ↳ publicații tipărite și materiale audiovizuale;
 - ↳ documente nepublicate, precum manuscrisele sau înregistrări audio și vizuale.
- ↳ Descrierea bibliografică a publicațiilor:
 - ↳ crearea înregistrărilor bibliografice pentru publicațiile din țară;
 - ↳ accesul utilizatorilor la înregistrările bibliografice, atât din țară cât și din alte țări.
- ↳ Disponibilitatea documentelor: obținerea, de oriunde din lume, a publicațiilor necesare pentru a satisface nevoia de informare a utilizatorilor.
- ↳ Accesul la documente:
 - ↳ acces pentru consultare în sediu, inclusiv electronic;
 - ↳ furnizare la distanță de reproducere sau împrumut.
- ↳ Accesul la informații independente de documente:
 - ↳ informații primare;
 - ↳ informații prelucrate, potrivit interesului utilizatorilor;
 - ↳ ghiduri de informare.
- ↳ Schimbul de publicații: eliminarea surplusului de material, prin redistribuire la și între biblioteci.
- ↳ Serviciile pentru alte biblioteci sau unități informare.
- ↳ Potențialul de dezvoltare:

- ↳ instruirea și formarea profesională;
- ↳ cercetarea și dezvoltarea.

3.2.1.2 Procese biblioteconomice

Principalele procese biblioteconomice se grupează pe trei domenii funcționale [36]:

↳ Dezvoltarea colecțiilor:

- ↳ elaborarea politicii de selecție a publicațiilor;
- ↳ identificarea surselor de unde pot fi obținute;
- ↳ selecția și achiziția publicațiilor.

↳ Organizarea colecțiilor:

- ↳ descrierea bibliografică a publicațiilor și constituirea cataloagelor de bibliotecă;
- ↳ aranjarea în depozite și conservarea publicațiilor;
- ↳ evaluarea colecțiilor și eliminarea de publicații.

↳ Utilizarea colecțiilor:

- ↳ împrumutul documentelor;
- ↳ asistența și îndrumarea utilizatorilor;
- ↳ elaborarea produselor de informare;

3.2.1.2 Măsurarea activităților

Procesele biblioteconomice sunt văzute ca succesiuni de activități formate la rândul lor din secvențe de operații elementare consumatoare de resurse.

Definiția 7. *Indicele de selecție* al unei operații, $s(\theta) \in \{0, 1\}$, este o valoare scalară care descrie faptul că, în conformitate cu o anumită politică de selecție (sau planificare), pentru o anumită operație analizată θ , descrisă într-un anumit nomenclator de operații $\theta \in \Theta$, se consideră necesară efectuarea ei:

$$s(\theta, \Theta) = \begin{cases} 1 & \text{dacă realizarea operației } \theta \text{ este considerată necesară;} \\ 0 & \text{în caz contrar.} \end{cases}$$

Definiția 8. *Indicele de realizare* al unei operații, $r(\theta) \geq 0$, este o valoare scalară care descrie faptul că o anumită operație analizată θ , descrisă într-un anumit nomenclator de operații $\theta \in \Theta$, a fost realizată:

$$r(\theta, \Theta) = \begin{cases} > 0 & \text{dacă operația } \theta \text{ a fost realizată;} \\ 0 & \text{în caz contrar.} \end{cases}$$

Observație. In procesele biblioteconomice curente se realizează și operații care nu fac obiectul unei politici de selecție.

3.2.1.3 Indicatori operaționali

Prin agregări ale indicilor de selecție sau de realizare se pot obține valorile indicatorilor care descriu starea curentă a sistemului bibliotecii [16], pe principalele domenii funcționale:

1. Dezvoltarea colecțiilor:

1.1 numărul total de volume Cărți și periodice tipărite (legate):

- ↗ numărul de volume intrate în timpul anului de referință;
- ↗ numărul de volume existente la sfârșitul anului de referință;
(cu evidențierea cărților și periodicelor rare, unități fizice/volume)

1.2 numărul total de unități fizice Manuscrise:

- ↗ numărul de unități fizice intrate în timpul anului de referință;
- ↗ numărul de unități fizice existente la sfârșitul anului de referință.

1.3 numărul total de unități fizice Microformate:

- ↗ numărul de unități fizice intrate în timpul anului de referință;
- ↗ numărul de unități fizice existente la sfârșitul anului de referință.

1.4 numărul total de unități fizice Documente audiovizuale, din care:

- ↳ documente audio;
- ↳ documente vizuale;
- ↳ documente audiovizuale combinate;
- ↗ numărul de unități fizice intrate în timpul anului de referință;
- ↗ numărul de unități fizice existente la sfârșitul anului de referință.

1.5 numărul total de unități fizice Colecții electronice, din care:

- ↳ documente digitale;
- ↳ periodice electronice;
- ↳ baze de date;
- (cu evidențierea tipului de suport: CD-ROM; DVD-ROM; WEB; ALT-TIP)
- ↗ numărul de unități fizice intrate în timpul anului de referință;
- ↗ numărul de unități fizice existente la sfârșitul anului de referință;

1.6 numărul total de unități fizice Alte documente de bibliotecă:

- ↗ numărul de unități fizice intrate în timpul anului de referință;
- ↗ numărul de unități fizice existente la sfârșitul anului de referință.

1.7 numărul total de titluri Cărți tipărite și electronice:

- ↗ numărul total de titluri intrate în timpul anului de referință;
- ↗ numărul total de titluri existente la sfârșitul anului de referință.

(Pentru indicatorii 1.1 – 1.7 se evidențiază numărul de documente în limba română, în alte limbi sau bilingve)

1.8 numărul total de Periodice curente (tipărite și electronice):

- ↗ numărul de titluri intrate în timpul anului de referință;
- ↗ numărul de abonamente în timpul anului de referință.

1.9 numărul de Resurse gratuite de pe Internet în timpul anului de referință.

1.10 numărul total de Documente achiziționate în timpul anului de referință, din care:

- ↗ din finanțare publică;
- ↗ din finanțare instituțională;

- ↖ din *venituri proprii*;
- ↖ din *alte venituri*.

1.11 numărul total de Documente eliminate în timpul anului de referință, din care:

- ↖ *cărți și periodice tipărite (legate) – volume*;
- ↖ *manuscrise – unități fizice*;
- ↖ *microformate – unități fizice*;
- ↖ *documente audiovizuale – unități fizice*;
- ↖ *colecții electronice – unități fizice*;
- ↖ *alte documente de bibliotecă – unități fizice*.

2. Accesul și facilitățile bibliotecii în timpul anului de referință:

2.1 numărul total de unități fizice Documente cu acces liber la raft.

2.2 numărul total de metri liniari de rafturi ocupate de colecții, din care:

- ↖ *numărul de metri liniari de rafturi cu acces liber ocupate de colecții*.

2.3 numărul total de înregistrări catalografice în sistem automatizat, din care:

- ↖ *numărul de documente rare înregistrate catalografic în sistem*.

2.4 numărul total de Calculatoare (PC).

2.5 numărul total de Stații de lucru pentru utilizatori, din care:

- ↖ *numărul de stații conectate la Internet*;
- ↖ *numărul de stații conectate la Intranet*.

2.6 numărul total de Imprimante.

2.7 numărul total de Scannere.

2.8 numărul total de Cititoare de cărți electronice.

2.9 numărul total de Fotocopitoare pentru utilizatori.

2.10 numărul total de Locuri în sălile de lectură.

2.11 numărul total de Sesiuni de comunicări științifice, simpozioane și conferințe.

2.12 numărul total de Programe și proiecte educative pentru populație – total din care:

- ↖ *pentru persoanele din grupa de vârstă: sub 15 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 15-24 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 25-64 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 65 ani și peste*.

2.13 numărul total de Cursuri de formare continuă.

2.14 numărul total de Activități în cadrul programelor educative – total din care:

- ↖ *pentru persoanele din grupa de vârstă: sub 15 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 15-24 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 25-64 ani*;
- ↖ *pentru persoanele din grupa de vârstă: 65 ani și peste*.

3. Utilizarea și utilizatorii bibliotecii în anul de referință:

3.1 numărul total de Utilizatori înscriși.

3.2 numărul total de Utilizatori activi, din care:

- ↖ după continuitate:

- ↳ *noi înscriși*;

- ↳ *vizați*.

- ↖ după naționalitate:

- ↳ *română*;

- ↳ *alte naționalități*.

- ↖ după gen:

- ↳ *masculin*;

- ↳ *feminin.*
- ↳ după vârstă:
 - ↳ *persoanele din grupa de vârstă: sub 15 ani;*
 - ↳ *persoanele din grupa de vârstă: 15-24 ani;*
 - ↳ *persoanele din grupa de vârstă: 25-64 ani;*
 - ↳ *persoanele din grupa de vârstă: 65 ani și peste.*
- ↳ după statutul ocupațional:
 - ↳ *profesii intelectuale;*
 - ↳ *tehnicieni;*
 - ↳ *funcționari;*
 - ↳ *muncitori;*
 - ↳ *elevi;*
 - ↳ *studenți;*
 - ↳ *pensionari;*
 - ↳ *casnice;*
 - ↳ *șomeri;*
 - ↳ *alte categorii.*

3.3 numărul total de Tranzacții de împrumut individual (documente eliberate), din care:

- ↳ după tipul documentelor
 - ↳ *cărți (volume);*
 - ↳ *periodice (volume);*
 - ↳ *fotocopii, microformate (unități fizice).*
- ↳ după conținutul documentelor:
 - ↳ *generalități;*
 - ↳ *filozofie, psihologie;*
 - ↳ *științe sociale, religie;*
 - ↳ *științe politice;*
 - ↳ *științe economice;*
 - ↳ *drept, administrație;*
 - ↳ *învățământ;*
 - ↳ *etnografie, folclor;*
 - ↳ *matematică, fizică, chimie;*
 - ↳ *geologie, biologie, botanică, zoologie;*
 - ↳ *medicină;*
 - ↳ *tehnică, inginerie, industrii;*
 - ↳ *comerț, marketing, management;*
 - ↳ *arte;*
 - ↳ *sport;*
 - ↳ *lingvistic;*
 - ↳ *literatură română;*

<ul style="list-style-type: none">↳ <i>literatură universală;</i>↳ <i>geografie;</i>↳ <i>arheologie, istorie, științe ale istoriei, biografii.</i> <p>3.4 numărul total de Titluri solicitate prin împrumut interbibliotecar</p> <ul style="list-style-type: none">↳ <i>Titluri solicitate de bibliotecă altor biblioteci, din care:</i><ul style="list-style-type: none">↳ <i>din țară;</i>↳ <i>din străinătate.</i>↳ <i>Titluri primite de bibliotecă de la alte biblioteci, din care:</i><ul style="list-style-type: none">↳ <i>din țară;</i>↳ <i>din străinătate.</i>↳ <i>Titluri solicitate bibliotecii de alte biblioteci, din care:</i><ul style="list-style-type: none">↳ <i>din țară;</i>↳ <i>din străinătate.</i>↳ <i>Titluri furnizate de bibliotecă altor biblioteci, din care:</i><ul style="list-style-type: none">↳ <i>din țară;</i>↳ <i>din străinătate.</i> <p>3.6 numărul total de Tranzacții de referințe virtuale.</p> <p>3.7 numărul total de Vizite la bibliotecă.</p> <p>3.8 numărul total de Vizite virtuale la resursele din rețea ale bibliotecii.</p> <p>3.9 numărul total de Participanți la Programe și proiecte educative pentru populație – din care:</p> <ul style="list-style-type: none">↳ <i>pentru persoanele din grupa de vârstă: sub 15 ani;</i>↳ <i>pentru persoanele din grupa de vârstă: 15-24 ani;</i>↳ <i>pentru persoanele din grupa de vârstă: 25-64 ani;</i>↳ <i>pentru persoanele din grupa de vârstă: 65 ani și peste.</i> <p>3.10 numărul total de Participanți la sesiuni de comunicări științifice, simpozioane și conferințe</p> <p>3.11 numărul total de Participanți la cursuri de formare continuă</p>
--

Tabelul 1. Lista indicatorilor operaționali ai bibliotecii

3.2.1.4 Indicatori de performanță

Managementul performant al unei instituții se bazează pe un set de obiective fixate și un sistem de indicatori de performanță asociați, menite să faciliteze sesizarea situațiilor decizionale, identificarea acțiunilor posibile de urmat și luarea deciziilor.

În continuare sunt prezentate obiectivele și indicatorii de performanță ai bibliotecii [17].

1. Dezvoltarea colecțiilor:

1.1 ponderea titlurilor deținute de bibliotecă în producția națională de publicații:

$$I_1 = (p_3 / P_3) \times 100, \text{ unde}$$

↳ p_3 := numărul de publicații achiziționate în ultimii 3 ani;

↳ P_3 := numărul total de publicații apărute în aceeași perioadă.

1.2 ponderea titlurilor deținute de bibliotecă în numărul total de titluri solicitate de către utilizatori:

$$I_2 = (s / S) \times 100, \text{ unde}$$

↪ s := numărul de publicații servite utilizatorilor în anul de referință;

↪ S := numărul total de publicații solicitate de utilizatori în același an.

2. Accesibilitatea colecțiilor:

2.1 ponderea publicațiilor din colecțiile generale, accesibile utilizatorilor în numărul total de publicații apărute în ultimii 2 ani:

$$I_3 = (p / P_2) \times 100, \text{ unde:}$$

↪ p := numărul de publicații catalogate (achiziționate și procesate) în anul de referință;

↪ P_2 := numărul total de publicații apărute în ultimii 2 ani.

Dobândirea și integrarea rapidă în colecțiile bibliotecii a titlurilor recent apărute este un criteriu major de eficiență. Procentajul mic poate avea multe cauze: procedurile pentru depozitul legal insuficient urmărite, lipsa preocupării pentru publicațiile noi, întârzieri în catalogare.

2.2 ponderea materialelor rare ce pot fi regăsite prin resurse web în numărul total de materiale din colecțiile speciale, reflectă gradul de preocupare al bibliotecii privind promovarea propriilor colecții:

$$I_4 = (m / M) \times 100, \text{ unde:}$$

↪ m := numărul de materiale rare accesibile pe web în anul de referință;

↪ M := numărul total de materiale rare existente în colecții, în același an.

3. Accesul rapid și ușor la publicații:

3.1 durata procesării publicațiilor – se referă la numărul mediu de zile dintre data intrării unei publicații în bibliotecă și apariția sa în inventar, catalog, resurse web etc., include achiziția, catalogarea descriptivă, indexarea, procesarea fizică.

Pentru fiecare publicație $i \in P$ se calculează $z(i)$ numărul de zile dintre data intrării și data de la care publicația a devenit accesibilă utilizatorilor.

$$I_5 = T / P, \text{ unde:}$$

↪ $T = \sum_{i=1}^P z(i)$:= numărul total de unități de timp (zile) consumate în anul de referință pentru catalogarea publicațiilor,

↪ P := numărul de publicații catalogate, în același an.

Scorul slab poate însemna comunicare insuficientă între compartimente și/sau acumulare de stocuri și/sau personal insuficient.

3.2 ponderea publicațiilor înregistrate care se află la locul corect:

$$I_6 = (p / P) \times 100, \text{ unde:}$$

↪ p := numărul de publicații corect așezate, în anul de referință;

↪ P := numărul total de publicații existente în același an.

3.3 durata medie a procesului de regăsire a unei publicații:

Pentru fiecare publicație $i \in I$ se calculează $r(i)$ numărul de unități de timp dintre momentul solicitării și momentul servirii.

$$I_7 = T / P, \text{ unde:}$$

↪ $T = \sum_{i=1}^P r(i)$:= numărul de unități de timp (ore) consumate pentru regăsirea publicațiilor, în anul de referință;

↪ P := numărul total al publicațiilor regăsite, în același an.

3.4 timpul mediu al schimburilor interbibliotecare – măsoară eficiența serviciilor de împrumut extern, se referă la timpul consumat pentru completarea formularelor necesare unui împrumut interbibliotecar:

$$I_8 = (T / F), \text{ unde}$$

↪ T := numărul total de unități de timp (zile) consumate în anul de referință pentru completarea formularelor de împrumut;

↪ F := numărul de formulare (de cerere și de primire) completate, în același an.

3.5 accesibilitatea directă a publicațiilor – descrie accesibilitatea celor mai solicitate materiale prin intermediul paginii web a instituției, fiind măsurată prin numărul de click-uri necesare pentru a se ajunge la documentul respectiv.

Pentru fiecare intrare web $i \in I$ se calculează $p(i)$ numărul de puncte acumulate.

$$I_9 = (b / B) \times 100, \text{ unde:}$$

↪ $b = (\sum_{i=1}^I p(i)) / I$:= numărul mediu de puncte acumulate de intrările web, în anul de referință;

↪ I := numărul total de intrări web, în același an;

↪ B := numărul total de butoane din meniu, în același an.

4. Utilizarea bibliotecii:

4.1 gradul de solicitare al publicațiilor străine:

$$I_{10} = (p_3 / P_3) \times 100, \text{ unde}$$

↪ p_3 := numărul total al împrumurilor efectuate în anul de referință cu publicații străine intrate în ultimii 3 ani;

↪ P_3 := numărul total al publicațiilor străine intrate în aceeași perioadă.

4.2 costul utilizării resurselor electronice:

$$I_{11} = C / U, \text{ unde:}$$

↪ C := costul resurselor electronice în anul de referință (baze de date, reviste electronice pentru care biblioteca plătește);

↪ U := numărul de unități de conținut descărcate, în același an.

4.3 gradul de ocupare a locurilor disponibile – proporția în care publicul utilizează resursele bibliotecii, valoarea bibliotecii ca spațiu fizic:

$$I_{12} = (l / L) \times 100, \text{ unde:}$$

↪ l := numărul mediu de locuri ocupate zilnic, în anul de referință;

↪ L := numărul total de locuri disponibile, în același an.

4.4 gradul de participare la evenimentele culturale – măsoară atractivitatea evenimentelor culturale organizate de bibliotecă:

$$I_{13} = P / E, \text{ unde:}$$

↪ P := numărul total de participanți la evenimente culturale, în anul de referință;

↪ E := numărul total de evenimente culturale organizate de bibliotecă, în același an.

4.5 gradul de satisfacere a utilizatorilor – exprimă aprecierea utilizatorilor privind serviciile oferite, ca un întreg sau ca servicii individuale. În general, pentru fiecare grup țintă de utilizatori există un chestionar special și o scară de valori numerice.

$$I_{14} = S / U, \text{ unde:}$$

↪ S := suma valorilor acordate de utilizatori, într-o anumită perioadă de referință;

↪ U := numărul total de utilizatori chestionați, în aceeași perioadă.

5. Digitalizarea colecțiilor:

5.1 gradul de digitizare a colecțiilor generale – măsoară gradul în care biblioteca își îndeplinește misiunea de a face accesibilă publicului moștenirea documentară în format digital:

$$I_{15} = (d / D) \times 1.000, \text{ unde:}$$

↪ d := numărul de documente digitizate în anul de referință;

↪ D := numărul total de documente, în același an.

5.2 gradul de digitizare a colecțiilor speciale – măsoară gradul în care biblioteca își îndeplinește misiunea de a face accesibilă publicului moștenirea documentară în format digital:

$$I_{16} = (d / D) \times 100, \text{ unde:}$$

↪ d := numărul de documente speciale digitizate în anul de referință;

↪ D := numărul total de documente în colecțiile speciale, în același an .

5.3 ponderea documentelor descărcate – descrie măsura în care biblioteca dispune de resurse digitale relevante necesare utilizatorilor:

$$I_{17} = (d / D) \times 100, \text{ unde:}$$

↪ d := numărul de documente downloadate din colecțiile digitale ale bibliotecii în anul de referință;

↪ D := numărul total de documente digitizate, în același an.

6. Servicii bibliografice, asistență, îndrumare

6.1 ponderea cererilor de asistență soluționate eficient în raport cu numărul total al cererilor de asistență:

$$I_{18} = (s / S) \times 100, \text{ unde:}$$

↪ s := numărul de solicitări de asistență soluționate corect în anul de referință;

↪ S := numărul total de solicitări de asistență, în același an.

Calificativele sunt următoarele :

↪ răspuns complet – răspuns corect și complet prin trimitere la o singură sursă sau trimitere la mai multe surse, dintre care una conține răspunsul corect și comoplet;

↪ răspuns parțial complet – trimitere la una sau mai multe surse, nici una dintre neconducând direct la răspuns, dar una cel puțin fiind în situația de a constitui sursă preliminară;

↪ referință – răspuns direct inexistent, dar trimitere la o altă persoană, instituție, sursă;

↪ incorect sau nici un răspuns – nici un răspuns, nici o indicație sau indicații complet greșite.

O valoare mare reflectă acuratețea mare a răspunsurilor. Poate furniza infirmații mai utile în cazul în care referințele sunt furnizate online. Poate fi influențat de natura întrebărilor, abilitatea personalului, calitatea, varietatea, disponibilitatea surselor și bazelor de date existente.

6.2 timpul mediu de răspuns la solicitările de asistență – descrie timpul mediu necesar personalului bibliotecii pentru a răspunde unei solicitări de asistență, măsurat în unități de timp (minute, ore, zile) în funcție de specificul fiecărui serviciu și măsoară gradul în care răspunsul se încadrează într-un interval de timp rezonabil.

$$I_{19} = (T / S), \text{ unde:}$$

↪ T := numărul total de unități de timp consumate în anul de referință pentru acordarea de răspunsuri la solicitările de asistență;

↳ S := numărul total de solicitări de asistență, în același an.

O valoare mică este considerată pozitivă, reflectând buna organizare a serviciilor bibliotecii.

7. Creșterea potențialului de dezvoltare.

7.1 timpul mediu alocat activităților de IT, pregătirea informatică – descrie importanța acordată de instituție implicării resurselor umane în serviciile electronice, gradul de prioritate acordat acestui aspect de către instituție. Indicatorul se referă la volumul timpului de lucru pe care personalul implicat îl dedică organizării, menținerii, oferirii și dezvoltării de resurse IT, dezvoltării tehnice și îmbunătățirii serviciilor electronice ale instituției:

$$I_{20} = (T / A), \text{ unde:}$$

↳ T := numărul total de unități de timp consumate în anul de referință de angajații bibliotecii pentru activități legate de IT;

↳ A := numărul total de angajați, în același an.

O valoare mare reflectă gradul mare de interes acordat de instituție dezvoltării.

7.2 ponderea pregătirii profesionale – măsoară importanța acordată de instituție pregătirii personalului propriu pentru creșterea abilităților profesionale, prin participări la cursuri de specializare.

$$I_{21} = (a / A) \times 100, \text{ unde:}$$

↳ a := numărul de angajați care au urmat cursuri de specializare în anul de referință;

↳ A := numărul total de angajați, în același an.

O valoare mare e considerată pozitivă, una mică reflectă necesitatea creșterii gradului de implicare a personalului în activități de instruire. Poate fi influențat de numărul și calitatea cursurilor oferite. Instituția poate promova cursuri prin rețeaua intranet sau încercând să afle necesitățile personalului pentru o pregătire absolut specifică, ca să iasă în întâmpinarea lor.

7.3 ponderea bunurilor obținute prin granturi – măsoară ponderea mijloacelor dobândite urmare a derulării unor granturi:

$$I_{22} = (v / V) \times 100, \text{ unde}$$

↳ v := valoarea bunurilor dobândite în anul de referință prin granturi sau resurse extrabugetare;

↳ V := valoarea totală a bunurilor instituției, în același an.

Obiectivul său este acela de a măsura performanța instituției în a obține resurse financiare adiționale și de aici abilitatea sa în procesul propriei dezvoltări. O valoare mare indică succesul instituției în a obține prin inițiativă proprie mijloace suplimentare. Măsoară și capacitatea bibliotecii de a se implica în sarcini care trec dincolo de sarcina sa principală și o califică pentru obținerea de mijloace suplimentare.

7.4 ponderea personalului implicat în granturi – măsoară intensitatea și spectrul cooperării la nivel național sau internațional al instituției și de aici importanța și impactul

instituției:

$$I_{23} = (a / A) \times 100, \text{ unde:}$$

- ↪ a := numărul de angajați implicați în proiecte și cooperări naționale sau internaționale în anul de referință;
- ↪ A := numărul total de angajați, în același an.

Un procentaj mare e considerat pozitiv, reflectând angajamentul instituției în provocările societății informaționale. Performanțele pot varia mult de la un an la altul, fiind influențate mult de disponibilitatea fondurilor. Utilitatea sa se vede mai ales dacă este aplicat consecutiv mulți ani.

8. Conservarea colecțiilor

8.1 gradul de accesibilitate a pieselor originale din colecții, măsoară capacitatea bibliotecii de conserva materialele originale, preocuparea ei pentru conservare și gradul de prioritate al acestei activități.

$$I_{24} = (e / E) \times 100, \text{ unde:}$$

- ↪ e := numărul de exemplare aflate în condiție stabilă, în anul de referință;
- ↪ E := numărul total de exemplare existente, în același an .

Obiectele sunt clasificate în patru categorii :

- ↪ aflate în condiție bună – utilizabile cu grija normală pentru orice material din colecție;
- ↪ condiție acceptabilă, corectă – deteriorat, dar stabil și utilizabil cu o precauție sporită;
- ↪ condiție precară – deteriorare moderată, imposibil de utilizat fără creșterea gradului de deteriorare;
- ↪ inutilizabil – grav deteriorat, exclus de la acces public și de la utilizare din cauza fragilității, infestării etc.

Primele două categorii pot fi considerate separat ca stabile, celelalte două ca instabile.

Indicatorul definește ponderea materialelor din colecție aflate în stare stabilă, adică apte a fi folosite fără riscul unor deteriorări. Dacă rezultatul arată un grad redus de stabilitate a condiției de conservare a materialelor, atunci trebuie luate măsuri specifice fiecărei categorii (manuscrise medievale, cartografie, monede) în vederea creșterea gradului de conservare și a prioritizării acestor activități. Măsurile țin de îmbunătățirea condițiilor de microclimat, schimbarea metodelor de utilizare și mănuire, tratament în vederea conservării, deacidificare, efectuarea de copii pentru public.

8.2 ponderea materialelor care au beneficiat de restaurare în raport cu volumul total al materialelor de restaurat:

$$I_{25} = (m / M) \times 100, \text{ unde:}$$

- ↪ m := numărul de materiale care au beneficiat de restaurare în anul de referință;
- ↪ M := numărul total de materiale care au nevoie de restaurare, în același an.

Materialele luate în calcul sunt cărțile rare publicate înainte de 1800, incunabulele și manuscrisele [18]. Un procentaj ridicat va fi considerat bun. Poate fi aplicat și pe categorii mai restrânse de material (manuscrise, ziare, hârti, stampe, monede). Dacă rezultatul este slab, atunci trebuie luate măsuri de promovare a importanței prezervării materialelor originale prin toate tipurile de media, încercarea de a obține fonduri suplimentare inclusiv prin tehnici de tipul *adoptați-o-carte* (obiect). Aceste tehnici ar putea avea succes având în vedere impactul public al materialelor rare.

8.3. ponderea spațiilor cu mediu ambiental adecvat – descrie în ce măsură colecțiile sunt asigurate cu un spațiu de depozitare adecvat și au o utilizare adecvată. Datele legate de temperatură și RH trebuie stocate un an de zile prin măsurători permanente cu aparatele necesare.

$$I_{26} = (s / S) \times 100, \text{ unde:}$$

- ↪ s := suprafața spațiilor de depozitare care beneficiază de condiții de mediu adecvate în anul de referință;
- ↪ S := suprafața totală a spațiilor de depozitare, în același an.

O valoare mare este considerată pozitivă. Trebuie utilizate filtre de ultraviolete, sticlă protectoare, draperii care să împiedice accesul luminii naturale.

9. Management eficient:

9.1 costul mediu al serviciilor de catalogare – măsoară costurile de personal pentru descrierea unui document și inserarea sa logică și coerentă într-un catalog. Scopul său este acela de a urmări costurile de personal și a le acomoda cu o politică specifică ce vizează generarea de înregistrări bibliografice.

$$I_{27} = (T \times h + C) / P, \text{ unde:}$$

- ↪ T := numărul total de ore consumate în anul de referință pentru înregistrări bibliografice, descrieri, identificări;
- ↪ h := costul unei ore de lucru;
- ↪ C := costul subcontractărilor, dacă este cazul;
- ↪ P := numărul de publicații catalogate, în același an.

Dacă rezultatul indică costuri mari, atunci pot fi luate măsuri precum: reconsiderarea complexității descrierii pentru diferite tipuri de documente; evaluarea performanțelor software și hardware; revizuirea fluxurilor de lucru; creșterea volumului de date copiate de la alte biblioteci.

9.2 costul mediu al serviciilor de împrumut – măsoară costurile pentru asigurarea serviciilor de împrumut:

$$I_{28} = (T \times h) / P, \text{ unde:}$$

- ↪ T := numărul de ore consumate în anul de referință pentru activitățile referitoare la împrumuturi;

↳ h := costul unei ore de lucru;

↳ P := numărul total de publicații împrumutate, în același an.

Dacă costurile de personal par a fi prea mari, măsurile de luat pot fi: computerizarea proceselor, revizuirea fluxurilor de lucru, înlocuirea personalului de specialitate, calificat, în anumite servicii, cu personal necalificat care poate îndeplini sarcini adecvate.

9.3 productivitatea serviciilor de catalogare – măsoară productivitatea activităților de procesare (achiziție și catalogare) a publicațiilor:

$$I_{29} = P / A, \text{ unde:}$$

↳ P := numărul total de publicații achiziționate și procesate în anul de referință;

↳ A := numărul de angajați implicați în achiziția și procesarea lor, în același an.

Rezultatul poate fi influențat de tipul de media supusă procesării, metodele de procesare, calificarea personalului. Dacă eficiența este redusă, instituția poate recurge la reconsiderarea nivelului de descriere a materialului, la revizuirea fluxurilor de lucru, la utilizarea mai largă a procedurilor automatizate, la pregătirea sporită a personalului sau la creșterea volumului de date importate de la alte biblioteci.

9.4 productivitatea serviciilor de împrumut – măsoară productivitatea activităților de împrumut, luând în calcul raportul dintre numărul anual al tranzacțiilor de împrumut, de toate felurile și numărul angajaților implicați.

$$I_{30} = P / A, \text{ unde:}$$

↳ P := numărul total de publicații împrumutate în anul de referință (în instituție, în afara ei, interbibliotecar);

↳ A := numărul de angajați implicați în activitățile de împrumut, în același an.

Valorile mari sunt considerate pozitive. Performanța poate fi influențată de tipul de material care urmează a fi livrat, ponderea diferitelor tipuri de împrumuturi, gradul de automatizare și fluxurile de lucru. Instituția poate recurge la reorganizarea fluxurilor de lucru, automatizarea sporită a proceselor, intensificarea pregătirii personalului.

Tabelul 2. Obiectivele bibliotecii și indicatorii de performanță

3.2.2 Cerințele bibliografice

3.2.2.1 Cerințe funcționale pentru datele bibliografice

Modelul conceptual de referință pentru datele bibliografice [4, 15] este modelul *FRBR* (*Functional Requirements for Bibliographic Records*) elaborat între 1992 și 1997 de către un grup de experți *IFLA* (*International Federation of Library Associations and Institutions*).

Modelul FRBR are o dublă orientare:

↳ Consolidarea acumulărilor trecute privind atât raționalizarea funcțiilor unui catalog de bibliotecă cât și justificarea științifică *à posteriori* a practicilor (costurilor) de catalogare la nivel internațional prin reconfirmarea valabilității *ISBD-urilor*

(*International Standard for Bibliographic Description*) și prin redefinirea informațiilor minime pe care trebuie să le conțină o înregistrare bibliografică;

- ↳ Oferirea de deschideri spre viitor stabilind fundamentele pentru cataloage inovatoare și asigurând cadrul conceptual și terminologic pentru actualizarea *Principiilor internaționale de catalogare* și pentru elaborarea *Codului internațional de catalogare (RDA - Resource Description and Access)*.

Deoarece *FRBR* s-a referit numai la informațiile bibliografice privind documentele, *IFLA* a extins eforturile de modelare la toate informațiile prezente într-un catalog de bibliotecă și în consecință:

- ↳ În aprilie 1999 a fost stabilit grupul de experți *FRANAR (Functional Requirements and Numbering of Authority Records)* cu misiunea de a modela conținutul înregistrărilor de autoritate. Grupul *FRANAR* a publicat, în 2009, modelul *FRAD (Functional Requirements for Authority Data)*.
- ↳ În aprilie 2005 a fost stabilit grupul de experți *FRSAR (Functional Requirements for Subject Authority Records)* responsabil pentru modelarea relațiilor dintre datele bibliografice și fișierul de autoritate discipline. Grupul *FRSAR* a finalizat în 2010 modelul *FRSAD (Functional Requirements for Subject Authority Data)*.
- ↳ Etapa ulterioară de dezvoltare constă în fuzionarea celor trei modele *FRBR*, *FRAD* și *FRSAD* într-unul singur.

Modelul *FRBR* a fost conceput în conformitate cu un formalism de tip *entitate-relație*. Dorința de a apropia acest model de modelul echivalent dezvoltat de comunitatea muzeografică, *CIDOC CRM (CIDOC Conceptual Reference Model)*, a condus ulterior la reformularea modelului *FRBR* în formalismul *orientat-obiect*, în care este exprimat acesta din urmă, cu consecința că în prezent există două versiuni distincte ale modelului: *FRBR_{ER}* și *FRBR_{oo}*.

Modelul *FRBR* consideră descrierea bibliografică a unui document ca pe o mulțime de informații privind patru aspecte diferite ale documentului descris:

- ↳ caracteristicile individuale ale unui exemplar al documentului;
- ↳ caracteristicile publicației de care aparține;
- ↳ caracteristicile conținutului intelectual sau artistic;
- ↳ caracteristicile creației abstracte la care se referă acest conținut (de exemplu: textul original al unui roman și o traducere a acestui roman se referă la o aceeași creație abstractă).

Aceste patru aspecte au fost denumite respectiv: *exemplar*, *manifestare*, *expresie* și *lucrare*.

La fiecare din aceste patru niveluri de analiză, documentul descris este pus în relație cu o *persoană* sau cu o *colectivitate* care a intervenit într-un mod specific la acest nivel.

Fiecare dintre aceste șase noțiuni, la care se mai adaugă încă alte patru *loc*, *eveniment*, *obiect* și *concept* pot constitui subiectele unei *lucrări*.

3.2.2.2 Definierea entităților și relațiilor

Entitățile definite în cadrul modelului *FRBR* reprezintă obiectele esențiale relevante pentru utilizatorii datelor bibliografice. Acestea sunt împărțite în trei grupuri:

- ↪ primul grup include *produse* ale unei activități intelectuale sau artistice care sunt numite sau descrise în înregistrările bibliografice: *lucrare*, *expresie*, *manifestare* și *exemplar*;
- ↪ al doilea grup include entități care au diferite *responsabilități* privind conținutul intelectual sau artistic, producția fizică, distribuția, gestionarea sau aspectele juridice ale unor astfel de produse: *persoană* și *comunitate*;
- ↪ al treilea grup include un alt set de entități care constituie *subiecte* ale demersului intelectual sau artistic: *concept*, *obiect*, *eveniment* și *loc*.

↳ *Produse* ale unei activități intelectuale sau artistice:

Identificator	Denumire	Definire	Comentariu
<i>EPw</i>	<i>lucrare</i>	o anumită creație/operă intelectuală sau artistică	entitate abstractă; identitatea de conținut a mai multor <i>expresii</i>
<i>EPe</i>	<i>expresie</i>	realizarea intelectuală sau artistică a unei <i>lucrări</i>	notație alfanumerică, muzicală sau coregrafică; formă sonoră, vizuală, obiectuală, cinetică sau combinații
<i>EPm</i>	<i>manifestare</i>	materializarea unei <i>expresii</i> a unei <i>lucrări</i>	în funcție de suport: manuscrise, cărți, periodice, afișe, fonograme, filme, casete, cd-uri
<i>EPI</i>	<i>exemplar</i>	un exemplar izolat al unei <i>manifestări</i>	un anumit exemplar al unei monografii, al unei casete audio, etc

Tabelul 3. Produsele unei activități intelectuale sau artistice

↳ *Responsabili* pentru conținutul intelectual sau artistic, realizarea materială și diseminarea sau pentru gestionarea și aspectele juridice, ale unei *lucrări*:

Identificator	Denumire	Definire	Comentariu
<i>ERp</i>	<i>persoană</i>	indivizi	autori, compozitori, artiști, editori, traducători, dirijori, interpreți, etc
<i>ERc</i>	<i>colectivitate</i>	organisme sau grupuri de indivizi și/sau colectivități	identificate prin un nume specific și care acționează ca un tot

Tabelul 4. Responsabilii pentru produsele unei activități intelectuale sau artistice

↳ *Subiecte* ale *lucrărilor*:

Identificator	Denumire	Definire	Comentariu
<i>ESc</i>	<i>concept</i>	o noțiune / idee abstractă	domenii de cunoaștere, discipline, teorii, metode, tehnici, practici, etc
<i>ESo</i>	<i>obiect</i>	o realitate materială	obiecte naturale sau artificiale, existente sau dispărute
<i>ESe</i>	<i>eveniment</i>	o acțiune sau un fapt	evenimente istorice, epoci, perioade cronologice
<i>ESp</i>	<i>loc</i>	date geografice/topografice	subiect al unei hărți, al unui atlas sau al unui ghid turistic etc

Tabelul 5. Subiectele produselor unei activități intelectuale sau artistice

Relațiile definite în cadrul modelului *FRBR* reprezintă legături esențiale între entitățile bibliografice, relevante pentru utilizatorii datelor bibliografice:

- ↳ O *lucrare* poate fi realizată prin una sau mai multe *expresii*; o *expresie* poate fi realizare a unei singure *lucrări*.
- ↳ O *expresie* poate fi materializată/incorporată în una sau mai multe *manifestări*; o *manifestare* poate materializa/incorpora una sau mai multe *expresii*.
- ↳ O *manifestare* poate fi reprezentată/exemplificată de unul sau mai multe *exemplare*; un *exemplar* poate reprezenta/exemplifica o singură *manifestare*.
- ↳ O *lucrare* poate fi creată de către una sau mai multe *persoane* și/sau *colectivități*; o *persoană* sau o *colectivitate* poate crea una sau mai multe *lucrări*.
- ↳ O *expresie* poate fi realizată de către una sau mai multe *persoane* și/sau *colectivități*; o *persoană* sau o *colectivitate* poate realiza una sau mai multe *expresii*.
- ↳ O *manifestare* poate fi produsă de către una sau mai multe *persoane* și/sau *colectivități*; o *persoană* sau o *colectivitate* poate produce una sau mai multe *manifestări*.
- ↳ Un *exemplar* poate fi deținut de către una sau mai multe *persoane* și/sau *colectivități*; o *persoană* sau o *colectivitate* poate deține unul sau mai multe *exemplare*.
- ↳ O *lucrare* poate avea drept *subiect* unul sau mai multe *concepte* și/sau *obiecte* și/sau *evenimente* și/sau *locuri*; un *concept* sau un *obiect* sau un *eveniment* sau un *loc* poate fi *subiect* ale uneia sau mai multor *lucrări*.
- ↳ O *lucrare* poate avea drept *subiect* una sau mai multe *lucrări* și/sau *expresii* și/sau *manifestări* și/sau *exemplare* și/sau *persoane* și/sau *colectivități*; o *lucrare* sau o *expresie* sau o *manifestare* sau un *exemplar* sau o *persoană* sau o *colectivitate* poate fi *subiect* ale uneia sau mai multor *lucrări*.
- ↳ O *lucrare* poate fi parte din o singură *lucrare*; o *lucrare* poate avea ca parte una sau mai multe *lucrări*.
- ↳ O *expresie* poate fi parte din o singură *expresie*; o *expresie* poate avea ca parte una sau mai multe *expresii*.
- ↳ O *manifestare* poate fi parte din o singură *manifestare*; o *manifestare* poate avea ca parte una sau mai multe *manifestări*.
- ↳ Un *exemplar* poate fi parte din un singur *exemplar*; un *exemplar* poate avea ca parte unul sau mai multe *exemplare*.
- ↳ O *lucrare* poate avea drept *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat* și/sau *adaptare* și/sau *transformare* și/sau *imitație* una sau mai multe *lucrări*; o *lucrare* poate fi *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat* și/sau *adaptare* și/sau *transformare* și/sau *imitație* pentru o singură *lucrare*.
- ↳ O *lucrare* poate avea drept *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat* și/sau *adaptare* și/sau *transformare* și/sau *imitație* una sau mai multe *expresii*; o *expresie* poate fi *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat* și/sau *adaptare* și/sau *transformare* și/sau *imitație* pentru o singură *lucrare*.
- ↳ O *expresie* poate avea drept *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat*

și/sau *adaptare* și/sau *transformare* și/sau *imitație* una sau mai multe *expresii*; o *expresie* poate fi *succesor* și/sau *supliment* și/sau *complement* și/sau *rezumat* și/sau *adaptare* și/sau *transformare* și/sau *imitație* pentru o singură *expresie*.

- ↳ O *expresie* poate avea drept *scurtare* sau *revizuire* sau *traducere* sau *aranjament* una sau mai multe *expresii*; o *expresie* poate fi *scurtare* sau *revizuire* sau *traducere* sau *aranjament* pentru o singură *expresie*.
- ↳ O *manifestare* poate fi *reproducere* pentru o singură *manifestare*; o *manifestare* poate avea drept *reproducere* una sau mai multe *manifestări*.
- ↳ Un *exemplar* poate fi *reproducere* pentru o singură *manifestare*; o *manifestare* poate avea drept *reproducere* unul sau mai multe *exemplare*.
- ↳ Un *exemplar* poate fi *reproducere* pentru un singur *exemplar*; un *exemplar* poate avea drept *reproducere* unul sau mai multe *exemplare*.
- ↳ O *manifestare* poate fi *alternativă* pentru o singură *manifestare*; o *manifestare* poate avea drept *alternativă* una sau mai multe *manifestări*.
- ↳ Un *exemplar* poate fi *reconfigurare* pentru un singur *exemplar*; un *exemplar* poate avea drept *reconfigurare* unul sau mai multe *exemplare*.

Tabelul 6. Relațiile dintre entitățile bibliografice

3.2.2.3 Descrierea entităților

↳ Atributele entității *EP_w* := **lucrare**

Identificator	Denumire atribut	Comentariu
<i>AP_w00</i>	id-w	
<i>AP_w01</i>	titlul <i>lucrării</i>	cuvânt, expresie sau grup de caractere
<i>AP_w02</i>	forma <i>lucrării</i>	roman, simfonie, hartă, pictură, fotografie, etc.
<i>AP_w03</i>	data <i>lucrării</i>	dată la care <i>lucrarea</i> a fost creată inițial
<i>AP_w04</i>	caracteristica distinctivă	diferențiază <i>lucrările</i> cu acelaș titlu
<i>AP_w05</i>	completitudinea vizată	indicație de finalizare/continuare nedefinită
<i>AP_w06</i>	publicul vizat	categorii de utilizatori cărora i se adresează
<i>AP_w07</i>	contextul <i>lucrării</i>	context istoric, social, intelectual, artistic etc.
<i>AP_w08</i>	distribuția executării	(<i>lucrare muzicală</i>)
<i>AP_w09</i>	referințe numerice	(<i>lucrare muzicală</i>)
<i>AP_w10</i>	tonalitate	(<i>lucrare muzicală</i>)
<i>AP_w11</i>	coordonate	(<i>lucrare cartografică</i>)
<i>AP_w12</i>	echinox	(<i>lucrare cartografică</i>)

↳ Atributele entității *E_{P_e}*

 := **expresie**

Identificator	Denumire atribut	Comentariu
<i>AP_e00</i>		
<i>AP_e01</i>	titlul <i>expresiei</i>	cuvânt, grup de cuvinte sau de caractere
<i>AP_e02</i>	forma <i>expresiei</i>	notație alfanumerică sau musicală, sunet,

		imagine, sculptură, dans, etc
APe03	data <i>expresiei</i>	data la care a fost creată <i>expresia</i>
APe04	limba <i>expresiei</i>	pot fi mai multe, pentru elemente diferite
APe05	caracteristică distinctivă	pentru a distinge <i>expresiile</i> aceleiași <i>lucrări</i>
APe06	aptitudinea de dezvoltare a <i>expresiei</i>	îmbogățirea în timp a conținutului
APe07	aptitudinea de revizuire a <i>expresiei</i>	versiune de lucru, actualizabil periodic, etc
APe08	volumul <i>expresiei</i>	număr de cuvinte, de imagini, durata, etc
APe09	rezumatul conținutului	un abstract, un sumar, un synopsis, un cuprins
APe10	contextul <i>expresiei</i>	contextul istoric, social, artistic, etc
APe11	receptarea <i>expresiei</i>	citarea opiniei unui critic, unui autor, etc
APe12	restricțiile de utilizare a <i>expresiei</i>	restricții bazate pe copyright
APe13	structurarea	(<i>périodic</i>) pe volume, numere, etc
APe14	regularitatea	(<i>periodic</i>) apariție regulată sau neregulată
APe15	periodicitatea	(<i>periodic</i>) lunar, trimestrial, anual, etc.
APe16	prezentarea muzicală	(notație muzicală)
APe17	distribuția de execuție	(notație muzicală sau înregistrare sonoră)
APe18	scara	(imagine/obiect cartografic)
APe19	proiecția	(imagine/obiect cartografic)
APe20	tehnica de prezentare	(imagine/obiect cartografic)
APe21	reprezentarea reliefului	(imagine/obiect cartografic)
APe22	măsurile geodezice, verticale; caroiaj	(imagine/obiect cartografic)
APe23	tehnica de înregistrare	(imagine de teledetecție)
APe24	caracteristica specială	(imagine de teledetecție)
APe25	tehnica	(imagine grafică sau proiectată)

↳ Atributele entității *EPm* := *manifestare*

Identificator	Denumire atribut	Comentariu
APm00		
APm01	titlul manifestării	unul sau mai multe titluri
APm02	mențiunea responsabilității	autor, interpret, editor, traducător
APm03	mențiunea ediției	ediția II , versiunea 3.0
APm04	locul ediției/difuzării	comună, oraș, provincie, țară
APm05	editorul/difuzorul	individ, grup, organism
APm06	data ediției/difuzării	data punerii la dispoziția publicului
APm07	fabricantul	individ, grup, organism
APm08	mențiunea colecției	indică numele colecției de care aparține
APm09	tipul de suport	casetă, microfilm, videodisc, diapozitiv
APm10	numărul de unitați materiale	foi (volante), discuri, bobine
APm11	materialul	hârtie, lemn, plastic, metal
APm12	modul de înregistrare	analogic, acustic, electric, numeric, optic, etc.
APm13	dimensiunile suportului	film de 8 mm, diapozitiv de 5 x 5 cm, plan în relief de 9 x 30 x 20 cm, disc de 30 cm
APm14	identificarea manifestării	ISBN, cotă, număr de inventar
APm15	furnizorul	nume și adresa pentru cel care autorizează accesul sau achiziția (editor, difuzor, etc.)
APm16	modalitățile de punere la dispoziție	gratuit, preț
APm17	restricțiile de acces la manifestare	bazate pe copyright,
APm18	stilul caracterelor	Baskerville, Times New Roman
APm19	mărimea caracterelor	10 puncte

<i>APm20</i>	formatul bibliografic	(cărți vechi) in-quarto, in-octavo, etc.
<i>APm21</i>	colaționarea	(cărți vechi) patru caiete notate de la A la D
<i>APm22</i>	starea publicației	(périodic) mort sau viu
<i>APm23</i>	numerotarea	(périodic) volumul 1, numărul 1, ianuarie 1971
<i>APm24</i>	viteza de rotație sau de defilare	(înregistrare sonoră)
<i>APm25</i>	lățimea canelurii	(înregistrare sonoră) nr canale/inch
<i>APm26</i>	tipul gravurii (direcția șanțurilor)	(înregistrare sonoră) lateral, vertical
<i>APm27</i>	configurația benzii	(înregistrare sonoră) numărul de piste
<i>APm28</i>	tipul de redare sonoră	(înregistrare sonoră) mono/stereo/quadro-fonic
<i>APm29</i>	caracteristicile speciale ale redării	(înregistrare sonoră)
<i>APm30</i>	culoarea	(imagine)
<i>APm31</i>	raportul de reducere	(microformă) 42x, de la 61x la 90x
<i>APm32</i>	polaritatea	(microformă/imagine proiectată)
<i>APm33</i>	generația	(microformă/imagine proiectată)
<i>APm34</i>	formatul de prezentare	(imagine proiectată) wide screen, Beta, VHS,
<i>APm35</i>	configurația necesară	(resursă electronică) hard, soft, periferice
<i>APm36</i>	caracteristicile fișierului	(resursă electronică) ASCII, SGML, etc.
<i>APm37</i>	modul de acces	(resursă electronique accesibilă la distanță)
<i>APm38</i>	adresa de acces (adresa URL)	(resursă electronique accesibilă la distanță)

↳ Atributele entității *E_{Pi}* := *exemplar*

Identificator	Denumire atribut	Comentariu
<i>APi00</i>		
<i>APi01</i>	identificarea <i>exemplarului</i>	cotă, număr de acces, cod debare, etc
<i>APi02</i>	amprenta	identificator, diferențiază exemplarele între ele
<i>APi03</i>	proveniența <i>exemplarului</i>	lista posesorilor/depozitarilor precedenți
<i>APi04</i>	adnotări/inscripționări	semnături, note, etc. unice
<i>APi05</i>	participările la expoziții	listă de expoziții publice, data, locul
<i>APi06</i>	starea <i>exemplarului</i>	starea fizică și/sau de conservare
<i>APi07</i>	operațiile de prezervare efectuate	procedurile, operațiile, datele intervențiilor
<i>APi08</i>	operațiile de prezervare de efectuat	procedurile, operațiile, datele intervențiilor
<i>APi09</i>	restricțiile de comunicare	orice limitare a accesului fizic la <i>exemplar</i>

↳ Atributele entității *E_{Rp}* := *persoană*

Identificator	Denumire atribut	Comentariu
<i>ARp00</i>		
<i>ARp01</i>	numele	numele <i>persoanei</i>
<i>ARp02</i>	datele	datele <i>persoanei</i>
<i>ARp03</i>	titlul	titlul <i>persoanei</i>
<i>ARp04</i>	altă caracteristică asociată	alt calificativ al <i>persoanei</i>

↳ Atributele entității *E_{Rc}* := *colectivitate*

Identificator	Denumire atribut	Comentariu
<i>ARc00</i>		
<i>ARc01</i>	numele	numele <i>colectivității</i>
<i>ARc02</i>	numărul asociat	numerotare pentru <i>colectivitate</i>

<i>ARc03</i>	locul	localizarea <i>colectivității</i>
<i>ARc04</i>	datele	datele asociate <i>colectivității</i>
<i>ARc05</i>	altă caracteristică asociată	alt calificativ al <i>colectivității</i>

↳ Atributele entității *ESc* := **concept**

Identificator	Denumire atribut	Comentariu
<i>ASc00</i>		
<i>ASc01</i>	numele <i>conceptului</i>	termen care denumește <i>conceptul</i>

↳ Atributele entității *ESo* := **obiect**

Identificator	Denumire atribut	Comentariu
<i>ASo00</i>		
<i>ASo01</i>	numele <i>obiectului</i>	termen care denumește <i>obiectul</i>

↳ Atributele entității *ESe* := **eveniment**

Identificator	Denumire atribut	Comentariu
<i>ASe00</i>		
<i>ASe01</i>	numele <i>evenimentului</i>	termen care denumește <i>evenimentul</i>

↳ Atributele entității *ESp* := **loc**

Identificator	Denumire atribut	Comentariu
<i>ASp00</i>		
<i>ASp01</i>	numele <i>locului</i>	termen care denumește <i>locul</i>

3.2.2.4 Descrierea relațiilor

↳ Relații $EP_w - EP_w$:= **lucrare – lucrare**

Identificator	Nume relație	Descriere	
<i>PwPw08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<— este <i>subiect</i> al
<i>PwPw09</i>	<i>Întreg/parte</i>	are ca <i>parte</i> —>>	<— este <i>parte</i> din
<i>PwPw10</i>	<i>Succesor</i>	are drept <i>succesor</i> —>>	<— este <i>succesor</i> al
<i>PwPw11</i>	<i>Supliment</i>	are drept <i>supliment</i> —>>	<— este <i>supliment</i> al
<i>PwPw12</i>	<i>Complement</i>	are drept <i>complement</i> —>>	<— este <i>complement</i> al
<i>PwPw13</i>	<i>Rezumat</i>	are drept <i>rezumat</i> —>>	<— este <i>rezumat</i> al
<i>PwPw14</i>	<i>Adaptare</i>	are drept <i>adaptare</i> —>>	<— este <i>adaptare</i> a
<i>PwPw15</i>	<i>Transformare</i>	are drept <i>transformare</i> —>>	<— este <i>transformare</i> a
<i>PwPw16</i>	<i>Imitație</i>	are drept <i>imitație</i> —>>	<— este <i>imitație</i> a

↳ Relații $EP_w - EP_e$:= **lucrare – expresie**

Identificator	Nume relație	Descriere	
<i>PwPe01</i>	<i>Realizare</i>	este <i>realizată</i> prin —>>	<— este <i>realizare</i> a
<i>PwPe08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<— este <i>subiect</i> al
<i>PwPe10</i>	<i>Succesor</i>	are drept <i>succesor</i> —>>	<— este <i>succesor</i> al

<i>PwPe11</i>	<i>Supliment</i>	are drept <i>supliment</i> —>>	<<— este <i>supliment</i> al
<i>PwPe12</i>	<i>Complement</i>	are drept <i>complement</i> —>>	<<— este <i>complement</i> al
<i>PwPe13</i>	<i>Rezumat</i>	are drept <i>rezumat</i> —>>	<<— este <i>rezumat</i> al
<i>PwPe14</i>	<i>Adaptare</i>	are drept <i>adaptare</i> —>>	<<— este <i>adaptare</i> a
<i>PwPe15</i>	<i>Transformare</i>	are drept <i>transformare</i> —>>	<<— este <i>transformare</i> a
<i>PwPe16</i>	<i>Imitație</i>	are drept <i>imitație</i> —>>	<<— este <i>imitație</i> a

↳ Relații $EP_w - EP_m := \text{lucrare} - \text{manifestare}$

Identificator	Nume relație	Descriere	
<i>PwPm08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - EP_i := \text{lucrare} - \text{exemplar}$

Identificator	Nume relație	Descriere	
<i>PwPi08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ER_p := \text{lucrare} - \text{persoană}$

Identificator	Nume relație	Descriere	
<i>PwRp04</i>	<i>Creație</i>	<i>creată de</i> —>>	<<— este <i>creator</i> al
<i>PwRp08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ER_c := \text{lucrare} - \text{colectivitate}$

Identificator	Nume relație	Descriere	
<i>PwRc04</i>	<i>Creație</i>	<i>creată de</i> —>>	<<— este <i>creator</i> al
<i>PwRc08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ESc := \text{lucrare} - \text{concept}$

Identificator	Nume relație	Descriere	
<i>PwSc08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ESo := \text{lucrare} - \text{obiect}$

Identificator	Nume relație	Descriere	
<i>PwSo08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ESe := \text{lucrare} - \text{eveniment}$

Identificator	Nume relație	Descriere	
<i>PwSe08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EP_w - ESp := \text{lucrare} - \text{loc}$

Identificator	Nume relație	Descriere	
<i>PwSp08</i>	<i>Subiect</i>	are drept <i>subiect</i> —>>	<<— este <i>subiect</i> al

↳ Relații $EPE - EPe := \text{expresie} - \text{expresie}$

Identificator	Nume relație	Descriere	
PePe09	Întreg/parte	are ca parte →>	← este parte din
PePe10	Succesor	are drept succesor →>	← este succesor al
PePe11	Supliment	are drept supliment →>	← este supliment al
PePe12	Complement	are drept complement →>	← este complement al
PePe13	Rezumat	are drept rezumat →>	← este rezumat al
PePe14	Adaptare	are drept adaptare →>	← este adaptare a
PePe15	Transformare	are drept transformare →>	← este transformare a
PePe16	Imitație	are drept imitație →>	← este imitație a
PePe17	Scurtare	are drept scurtare →>	← este scurtare a
PePe18	Revizuire	are drept revizuire →>	← este revizuire a
PePe19	Traducere	are drept traducere →>	← este traducere a
PePe20	Aranjament	are drept aranjament →>	← este aranjament al

↳ Relații $EPe - EPm := expresie - manifestare$

Identificator	Nume relație	Descriere	
PePm02	Materializare	este încorporată în →>	<< ← încorporează

↳ Relații $EPe - ERp := expresie - persoană$

Identificator	Nume relație	Descriere	
PeRp05	Realizare	realizată de →>	<< ← este realizator al

↳ Relații $EPe - ERc := expresie - colectivitate$

Identificator	Nume relație	Descriere	
PeRc05	Realizare	realizată de →>	<< ← este realizator al

↳ Relații $EPm - EPm := manifestare - manifestare$

Identificator	Nume relație	Descriere	
PmPm09	Întreg/parte	are ca parte →>	← este parte din
PmPm21	Reproducere	are drept reproducere →>	← este reproducere a
PmPm22	Alternativă	are drept alternativă →>	← este alternativă a

↳ Relații $EPm - EPi := manifestare - exemplar$

Identificator	Nume relație	Descriere	
PmPi03	Reprezentare	este exemplificată prin →>	← exemplifică
PmPi21	Reproducere	are drept reproducere →>	← este reproducere a

↳ Relații $EPm - ERp := manifestare - persoană$

Identificator	Nume relație	Descriere	
PmRp06	Producere	este produsă de →>	<< ← este producător al

↳ Relații $EPm - ERc := manifestare - colectivitate$

Identificator	Nume relație	Descriere	
<i>PmRc06</i>	<i>Producere</i>	este <i>produsă</i> de —>>	<<— este <i>producător</i> al

↪ Relații $EP_i - EP_i := \text{exemplar} - \text{exemplar}$

Identificator	Nume relație	Descriere	
<i>PiPi09</i>	<i>Întreg/parte</i>	are ca <i>parte</i> —>>	<— este <i>parte</i> din
<i>PiPi21</i>	<i>Reproducere</i>	are drept <i>reproducere</i> —>>	<— este <i>reproducere</i> a
<i>PiPi23</i>	<i>Reconfigurare</i>	are drept <i>reconfigurare</i> —>>	<— este <i>reconfigurare</i> a

↪ Relații $EP_i - ER_p := \text{exemplar} - \text{persoană}$

Identificator	Nume relație	Descriere	
<i>PiRp07</i>	<i>Posesie</i>	este <i>deținut</i> de —>>	<<— <i>deține</i>

↪ Relații $EP_i - ER_c := \text{exemplar} - \text{colectivitate}$

Identificator	Nume relație	Descriere	
<i>PiRc07</i>	<i>Posesie</i>	este <i>deținut</i> de —>>	<<— <i>deține</i>

3.2.3 Cerințele bibliometrice

3.2.3.1 Indicatori bibliometrici

Cercetătorii, ca autori, elaborează diverse tipuri de lucrări care conțin rezultate experimentale, teorii, recenzii etc.

Cea mai simplă metodă pentru a măsura productivitatea unui cercetător sau a unui grup de cercetare este de a număra documentele publicate (publicările) de un anumit autor sau de un anumit grup dintr-un anumit interval de timp. Indicatorii destinați pentru a măsura productivitatea cercetătorilor sau a grupurilor de cercetare sunt considerați indicatori cantitativi [8].

Pentru a raporta constatările lor la rezultate anterioare autorii citează alte lucrări. În mod firesc un autor citează un anumit document dacă:

- ↪ a studiat acel document;
- ↪ îl consideră relevant pentru lucrarea proprie;
- ↪ îl consideră suficient de important pentru a-l cita în mod explicit.

Modalitatea actuală de a măsura impactul lucrărilor științifice se bazează pe numărul de citări. Indicatorii care ajută la identificarea nivelului de calitate al lucrărilor unui cercetător sau ale unui grup de cercetare și pot fi utilizați pentru a evalua impactul cercetărilor în comunitatea științifică sunt considerați indicatori de performanță [8].

Descoperirile științifice și rezultatele cercetării sunt publicate în reviste științifice sau pe site-uri Web. Înainte de publicare articolele științifice sunt, în general, supuse controlului unor experți din domeniul respectiv (*peer review*). După publicare, rezultatele cercetărilor sunt utilizate de către alți cercetători pentru studiile lor și citate ca referințe în articolele ulterioare ale acestora. Citarea articolelor utilizate este caracteristică publicațiilor științifice și este

general acceptat faptul că numărul de citări ale unui anumit articol este o reflectare a impactului acesteia în comunitatea științifică.

Citările furnizează date care pot fi utilizate pentru a măsura statistic și matematic importanța relativă a unui articol sau a unei publicații precum și conectivitatea dintre domenii științifice, dintre departamente de cercetare sau dintre autori.

Aplicarea de metode matematice și statistice de analiză a cărților, a articolelor și a altor mijloace de comunicare, pentru a le măsura cantitatea și performanțele (de exemplu, calitatea) este cunoscută în lumea științifică sub numele de bibliometrie.

Ca sistem de măsurare a performanțelor producțiilor științifice, bibliometria, exprimată prin diferiți indicatori, a devenit foarte importantă atât pentru cercetători cât și pentru organizații.

Pentru cercetători, indicatorii bibliometrici sunt importanți deoarece permit măsurători obiective ale difuzării și a impactului în comunitatea științifică a articolelor publicate de o anumită revistă. Practic, acești indicatori pot ajuta cercetătorii în selectarea revistelor la care să-și prezinte manuscrisele lor.

Pentru organizații, indicatorii bibliometrici sunt importanți deoarece permit măsurători obiective privind calitatea unei anumite cercetări, unui cercetător individual sau a unui grup de cercetare. Practic, acești indicatori pot ajuta organizațiile să ia decizii privind angajarea, promovarea și finanțarea.

Mai mult decât atât, indicatorii bibliometrici pot fi utilizați pentru a determina originea geografică a cercetărilor și a detecta creșterea sau eroziunea impactului științific la nivel de țări. Ei pot fi folosiți pentru a măsura concentrarea cercetărilor și pentru a detecta alocarea greșită a resurselor de cercetare. Aceștia pot fi de asemenea folosiți pentru previzionarea succesului cercetării precum și fondurile de cercetare privind optimizarea politicii de cercetare.

Trei legi sunt frecvent utilizate în bibliometrie și anume legea productivității științifice a lui Lotka, legea dispersiei a lui Bradford și legea frecvenței cuvintelor a lui Zipf.

Legea lui Lotka descrie frecvența cu care publică autorii dintr-un anumit domeniu: *în mulțimea tuturor autorilor care publică într-un anumit domeniu, $A(c)$, procentul de autori care publică n articole este de $60 / n^2$.*

Legea lui Bradford servește bibliotecarilor ca un ghid general în determinarea numărului de reviste de bază din orice domeniu: *dacă $|P(c)|$ reprezintă numărul total de reviste din un anumit domeniu c și $|D(c)|$ reprezintă numărul total de articole publicate în aceste reviste atunci există constantele n și q astfel încât cele $|P(c)|$ reviste pot fi organizate în trei grupuri, fiecare grup de publicații publicând cam același număr de articole, $\approx |D(c)| / 3$ și $|P(c)| \approx (n \times q^0) + (n \times q^1) + (n \times q^2)$.*

Legea lui Zipf este utilizată pentru a prezice frecvența cuvintelor în cadrul unui text: *pentru un text destul de lung, dacă se listează cuvintele în ordinea descrescătoare a frecvenței lor, rangul unui cuvânt din această listă înmulțit cu frecvența sa va fi egal cu o constantă:*

$$r \times f = k, \text{ unde:}$$

r este rangul cuvântului, f este frecvența iar k este o constantă, $k \approx 26.500$ (Potter 1988).

3.2.3.2 Indicatori bibliometrici de productivitate

Pentru entitățile care fac obiectul analizelor bibliometrice sunt utilizate, în continuare, următoarele notații:

- ↪ $D \equiv$ mulțimea documentelor (publicărilor), $d \in D$;
- ↪ $A \equiv$ mulțimea autorilor, $a \in A$;
- ↪ $G \equiv$ mulțimea grupurilor de cercetare (centre de cercetare, universități, etc. la care sunt eventual afiliați autorii) $g \in G$;
- ↪ $P \equiv$ mulțimea publicațiilor, $p \in P$;
- ↪ $R \equiv$ mulțimea responsabililor de ediții (editorilor), $r \in R$;
- ↪ $E \equiv$ mulțimea editurilor, $e \in E$;
- ↪ $C \equiv$ mulțimea domeniilor de cercetare $c \in C$
- ↪ $S \equiv$ mulțimea subiectelor, ca părți ale unui (sub)domeniu de cercetare, $s \in S$;
- ↪ $x(y) \equiv$ entitate, din mulțimea X , care conține pe y ;
- ↪ $X(y) \equiv$ secțiune a mulțimii X definită de y , adică, toate entitățile din mulțimea X care conțin pe y ;
- ↪ $X(Y) \equiv \cup_{y \in Y} X(y)$;
- ↪ $x(y, z) \equiv$ entitate din mulțimea X , care conține și pe y și pe z
- ↪ $X(y, z) \equiv$ submulțimea entităților din $X(y)$ care conțin pe z . $X(y, z) = X(y) \cap X(z)$

Definiția 9. *Indicele de publicare* al unui document d , este o valoare scalară, $\pi(d, x')$, care descrie faptul că una sau mai multe entități $x' \in X(d)$, $X(d) \subset X$ au contribuit în mod specific la publicarea documentului $d \in D$:

$$\pi(d, x') = \begin{cases} 1 & \text{dacă } x' \in X(d) \\ 0 & \text{în caz contrar} \end{cases}$$

Definiția 10. *Indicele de cotare* al unei publicații p este o valoare scalară, $\psi(p, k')$, care descrie faptul că publicația p satisface un anumit criteriu de selecție k' (ISI, top journal ș.a.):

$$\psi(p, k') = \begin{cases} 1 & \text{dacă } p \text{ satisface criteriul } k' \\ 0 & \text{în caz contrar} \end{cases}$$

Definiția 11. *Indicele de cotare al unui document:* $\psi(d, k') = \psi(p, k')$ unde $\pi(d, p) = 1$.

Definiția 12. *Funcția de publicare* este o funcție $I(x, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(x, t)$ este o valoare scalară care reprezintă numărul tuturor publicațiilor produse de entitatea analizată, x , înainte de momentul t .

O variantă, simplă și calculabilă, de definire a funcției de publicare pentru entitatea analizată x și intervalul de timp analizat $T = [0, t-1]$ este:

$$I(x, t) = \sum_{d \in D(x, T)} \pi(d, x), \text{ unde}$$

$$\hookrightarrow D(x, T) = \{ d \mid d \in D(x), t(d) \in T \}.$$

Tabelul de mai jos conține majoritatea indicatorilor bibliometrici de productivitate specificați în literatura de specialitate [8, 31, 39]:

1.1 numărul total de publicări produse de entitatea analizată, x , în intervalul de timp analizat $T = [t', t'']$:

$$P \equiv \Pi(x, T) = \sum_{d \in D(x, T)} \pi(d, x).$$

1.2 numărul de publicări având co-autori dintr-o altă entitate x' produse de entitatea analizată x în intervalul de timp analizat $T = [t', t'']$:

$$P_{x'} = \sum_{a' \in A(x')} \sum_{d \in D(x, T)} \pi(d, x) \pi(d, a').$$

1.3 co-authoring – ponderea publicărilor având co-autori dintr-o altă entitate, se obține împărțind numărul de publicări cu co-autori dintr-o entitate selectată x' , la numărul total de publicări produse de unitatea analizată x în intervalul de timp analizat $T = [t', t'']$:

$$p_{x'} = P_{x'} / \Pi(x, T).$$

Acest indicator este utilizat pentru a arăta în ce măsură o entitate analizată cooperează cu alte entități în producerea de articole:

- ↪ colaborare internațională – ponderea publicărilor cu co-autori de la organizații din cel puțin două țări diferite;
- ↪ colaborare națională – ponderea publicărilor cu co-autori de la cel puțin două organizații din aceeași țară;
- ↪ colaborare departamentală – ponderea publicărilor cu co-autori de la cel puțin două departamente din cadrul aceleiași organizații;
- ↪ colaborare organizațională – ponderea publicărilor cu co-autori din două sau mai multe unități de cercetare.

1.4 number of publications in Thomson ISI indices – numărul total de publicări, cotate ISI, produse de unitatea analizată x în intervalul de timp analizat $T = [t', t'']$:

$$P_{ISI} = \sum_{d \in D(x, T)} \pi(d, x) \psi(d, ISI).$$

1.5 number of publications in top journals – numărul de publicări, în intervalul de timp analizat $T = [t', t'']$, ale unității analizate în publicații selectate cu un criteriu adecvat:

$$P_{TJ} = \sum_{d \in D(x, T)} \pi(d, x) \psi(d, TJ).$$

1.6 CEST field-based world share of publications – ponderea publicărilor $\Pi(x, c, T)$, din intervalul de timp analizat $T = [0, t-1]$ produse de entitatea analizată x în fiecare domeniu $c \in C(x)$ în care entitatea este activă, în totalul publicărilor din domeniul respectiv $\Pi(c, T)$:

$$p_w = 1000 (\sum_{c \in C(x)} \Pi(x, c, T) \Pi(c, T)) / (\sum_{c \in C(x)} \Pi(c, T)), \text{ unde}$$

$$\Psi(x, c, T) = \sum_{d \in D(c, T)} \pi(d, x);$$

$$\Psi(c, T) = \sum_{d \in D(c, T)} \pi(d, c)$$

1.7 CEST degree of specialization – gradul de specializare al unei entități analizate x pe intervalul de timp analizat $T = [0, t-1]$:

$$DS = (\sum_{c \in C} \square_c) / (|C| (100)^2), \text{ unde:}$$

$$\square_c = 100R_c - 100;$$

$$R_c = (\Pi(x, c, T) / \sum_{c \in C} \Pi(x, c, T)) / (\Pi(c, T) / \sum_{c \in C} \Pi(c, T));$$

↪ $c \in C$ și $|C| = 107$, numărul de domenii pentru publicațiile cotate *ISI*.

Gradul de specializare al unei entități analizate este un indicator structural bazat pe numărul de domenii de cercetare în care entitatea analizată este activă și pe numărul de publicări produse în fiecare din ele.

Gradul de specializare global este prin definiție 0. O entitate foarte specializată poate avea un grad maxim de specializare 1. Între aceste două extreme există 5 clase:

- ↪ grad foarte redus de specializare : $DS < 0,2$;
- ↪ grad redus de specializare : $0,2 \leq DS < 0,4$;
- ↪ grad mediu de specializare : $0,4 \leq DS < 0,6$;
- ↪ grad înalt de specializare : $0,6 \leq DS < 0,8$;
- ↪ grad foarte ridicat de specializare : $DS \geq 0,8$.

1.8 CEST relative activity index – indicatorul de activitate relativă descrie faptul că entitatea analizată x este mai mult sau mai puțin activă în domeniile de cercetare alese în raport cu restul lumii; valorile sunt normalizate într-un interval de la 0 la 200 în care 100 reprezintă media mondială:

$$RAI = 100 + 100 * (R_c^2 - 1) / (R_c^2 + 1), \text{ unde}$$

$$\hookrightarrow R_c = (\Pi(x, c, T) / \sum_{c \in C} \Pi(x, c, T)) / (\Pi(c, T) / \sum_{c \in C} \Pi(c, T)).$$

Tabelul 7. Indicatori bibliometrici de productivitate

3.2.3.3 Indicatori bibliometrici de performanță

Definiția 13. *Indicele de impact* este o valoare scalară, $\rho(d, d') \in \{0, 1\}$, care descrie faptul că un anumit document $d \in D$ a fost citat într-un alt document d' , $d' \in D$, $d' \neq d$:

$$\rho(d, d') = \begin{cases} 1 & \text{dacă pentru } d \text{ există o referință în } d' \\ 0 & \text{în caz contrar} \end{cases}$$

Dacă documentele $d' \neq d''$ citează amândouă un anumit document \vec{d} se spune că sunt cuplate bibliografic chiar dacă ele nu se citează direct unul pe altul. Cu cât citează împreună mai multe documente, cu atât relația lor este mai puternică.

Definiția 14. *Indicele de cuplare bibliografică* a două documente, $d' \neq d''$, este o valoare scalară, $\omega(d', d'') \in N$, care descrie faptul că articolele d' și d'' citează amândouă documentul $\vec{d} \in D$:

$$\omega(d', d'') = |\mathcal{D}|, \text{ unde } \mathcal{D} = \{ \vec{d} \in D \mid \rho(\vec{d}, d') = 1 \wedge \rho(\vec{d}, d'') = 1 \} \text{ sau}$$

$$\omega(d', d'') = \sum_{\vec{d} \in \mathcal{D}} \rho(\vec{d}, d') \rho(\vec{d}, d'').$$

Dacă două documente $d' \neq d''$ sunt citate împreună de către un anumit document \vec{d} se spune că sunt cuplate prin co-citare chiar dacă acestea nu se citează direct între ele. Cu cât sunt citate împreună de mai multe documente cu atât relația lor este mai puternică. Cuplarea co-citărilor este o metodă utilizată pentru a stabili o similitudine de subiect între două documente.

Definiția 15. *Indicele de cuplare prin co-citare* a două documente, $d' \neq d''$, este o valoare scalară, $w(d', d'') \in \mathbb{N}$, care descrie faptul că articolele d' și d'' sunt citate împreună de articolele $\vec{d} \in \mathcal{D} \subseteq D$:

$$w(d', d'') \equiv |\mathcal{D}|, \text{ unde } \mathcal{D} = \{\vec{d} \in D \mid (\rho(d', \vec{d}) = 1) \wedge (\rho(d'', \vec{d}) = 1)\} \text{ sau}$$

$$w(d', d'') = \sum_{\vec{d} \in \mathcal{D}} \rho(d', \vec{d}) \rho(d'', \vec{d}).$$

În marea majoritate a cazurilor, autorii documentelor extrem de citate au avut un impact major în domeniul lor de cercetare, dar există și multe situații în care numărul mare de citări nu susține un autor relevant. Din cauza practicilor de inginerie a citărilor, răspândite în publicare, autocitări, co-autori, cluburi de citare și alte tehnici pentru a obține rezultate umflate, astfel de citări devin evident mai puțin relevante. O soluție menită să diminueze erorile din sistemul actual poate fi obținută prin definirea unei măsuri a impactului lucrărilor științifice bazată atât pe *credibilitatea documentului* care citează cât și pe *relevanța citării* respective.

Definiția 16. *Indicele de notorietate al unei entități analizate*, x , este un scor $\varepsilon(x)$ atașat lui x de către experți, membri ai unor centre recunoscute ca autorități științifice.

Indicele de notorietate este o măsură a competenței într-un anumit (sub)domeniu de cercetare definit de (sau care include) un anumit subiect de interes. Indicele de notorietate este în funcție de anumite opinii personale deci este o măsură subiectivă a competenței.

Atribuirea scorurilor se bazează pe informații obținute prin:

- ↳ interviuri/sondaje ale unor experți;
- ↳ analiza unor bibliografii;
- ↳ pentru entitățile care nu au nici o referire în interviuri, sondaje sau bibliografii scorul atribuit implicit este zero.

Scorurile sunt înregistrate în liste de notorietate cu clasificări ale autorilor, publicațiilor, centrelor de cercetare, editurilor și documentelor. Entitățile fără nici o referire în interviuri, sondaje sau bibliografii nu sunt înregistrate în listele de notorietate.

În ceea ce privește documentele, de regulă, un document izolat d nu are atașat un indice de notorietate $\varepsilon(d)$, înregistrarea unui document într-un nomenclator de notorietate a documentelor este o excepție. Listele de documente de notorietate conțin exclusiv înregistrări pentru texte care aparțin unui „stoc de documente de bază” respectiv documente esențiale și de actualitate pentru un anumit domeniu de cercetare.

Definiția 17. *Indicele de încredere al unui document d* este un indice $\varepsilon(d)$, care depinde de toți sau de o parte a indicilor de notorietate atașați entităților care sunt în relație cu acel document, respectiv autorul, editura, publicația sau grupul la care este afiliat autorul:

$$\varepsilon(d) = \phi(\varepsilon(A(d)), \varepsilon(E(d)), \varepsilon(P(d)), \varepsilon(G(A(d)))) .$$

O variantă simplă, dar evaluabilă, de definire a indicelui de încredere pentru un document d este:

$$\varepsilon(d) = (w_A \varepsilon(A(d)) + w_E \varepsilon(E(d)) + w_P \varepsilon(P(d)) + w_G \varepsilon(G(A(d)))) / \mathcal{E}, \text{ unde}$$

$$\text{↳ } w_A + w_E + w_P + w_G = 1; w_A, w_E, w_P, w_G \geq 0;$$

$$\text{↳ } \mathcal{E} = \varepsilon(A(d)) + \varepsilon(E(d)) + \varepsilon(P(d)) + \varepsilon(G(A(d))).$$

$\varepsilon(d)$ este un indice a priori, care descrie un document d în momentul publicării, înainte de a se obține informații despre referințele la d .

Definiția 18. *Indicele de relevanță al unei citări* este o valoare scalară, $\sigma(d, d') \geq 0$, care descrie cât de relevantă poate fi considerată citarea documentului $d \in D$ de către documentul $d' \in D$:

$$\sigma(d, d') \begin{cases} > 0 & \text{dacă } d \text{ este citat în } d' \\ = 0 & \text{în caz contrar} \end{cases}$$

O formulă simplă, dar evaluabilă, pentru indicele de relevanță al unei citări este:

$$\sigma(d, d') = M / (m + M), \text{ unde:}$$

$$\rightsquigarrow d \in D(a), d' \in D(a'), M = \max\{\rho(a, a'), \rho(a', a)\} \text{ și } m = \min\{\rho(a, a'), \rho(a', a)\}.$$

Observații. $\sigma(d, d') \in [0.5, 1.0]$; $m \equiv$ număr de citări reciproce ($a \neq a'$) sau număr de autocitări ($a = a'$).

2.1 numărul total de citări ale unui anumit document, $d \in D$:

$$\rho(d) \equiv \rho(d, D) = \sum_{d' \in D} \rho(d, d')$$

2.2 numărul de citări ale unui document, $d \in D$, de către un autor, $a' \in A$:

$$\rho(d, a') \equiv \rho(d, D(a')) = \sum_{d' \in D(a')} \rho(d, d')$$

2.3 numărul total de citări ale unui autor, $a \in A$:

$$\rho(a) \equiv \rho(D(a), D) = \sum_{d \in D(a)} \sum_{d' \in D} \rho(d, d')$$

2.4 numărul de citări ale unui autor $a \in A$ de către alt autor $a' \in A$:

$$\rho(a, a') \equiv \rho(D(a), D(a')) = \sum_{d \in D(a)} \sum_{d' \in D(a')} \rho(d, d')$$

2.5 numărul de citări ale unui document, $d \in D$, de către o publicație, $p' \in P$:

$$\rho(d, p') \equiv \rho(d, D(p')) = \sum_{d' \in D(p')} \rho(d, d')$$

2.6 numărul de citări ale unei publicații, $p \in P$ de către o altă publicație, $p' \in P$:

$$\rho(p, p') \equiv \rho(D(p), D(p')) = \sum_{d \in D(p)} \sum_{d' \in D(p')} \rho(d, d')$$

2.7 numărul total de citări ale unei publicații, $p \in P$:

$$\rho(p) \equiv \rho(D(p), D) = \sum_{d \in D(p)} \sum_{d' \in D} \rho(d, d')$$

2.8 numărul total de citări primite de entitatea analizată, x , în intervalul de timp analizat, T :

$$\rho(x, T) \equiv \rho(D(x), D(T)) = \sum_{d \in D(x)} \sum_{d' \in D(T)} \rho(d, d')$$

2.9 numărul de citări primite de o entitate analizată, x , de la o entitate selectată, x' , în intervalul de timp analizat, T :

$$\rho(x, x', T) \equiv \rho(D(x), D(x', T)) = \sum_{d \in D(x)} \sum_{d' \in D(x', T)} \rho(d, d')$$

2.10 Hirsch index (*h-index*) – *indicele Hirsch* [14] reprezintă numărul de publicări produse de entitatea analizată x în perioada de timp analizată T care au cel puțin h citări, $\Pi_h(x, T) \subset \Pi(x, T)$. Articolele publicate de entitatea analizată sunt sortate în ordinea descrescătoare a numărului de citări, se numără articolele începând din partea superioară a listei și atunci când numărul curent al unui articol depășește numărul de citări al acelui articol, numărul curent al articolului precedent este luat în considerare ca *h-index*.

$$h \equiv h\text{-index}(x, T) = i, \text{ unde:}$$

$$\Leftrightarrow [i \leq \rho(d'_i)] \wedge [i+1 > \rho(d'_{i+1})]; d'_i \in D(x, T) \text{ și}$$

$$\Leftrightarrow D(x, T) = \{ d'_i \mid d'_i \in D(x, T), \rho(d'_{i+1}) \geq \rho(d'_i), (\forall) i \in [1, \Pi(x, T) - 1] \}.$$

2.11 cited half-life – funcția de înjumătățire a numărului de citări ale unei publicații p , este o funcție $H_1(p, \bullet) : P \rightarrow \mathfrak{R}^+$ unde $H_1(p, t)$ este o valoare scalară care reprezintă intervalul de timp (numărul de ani), anterior anului t , care acoperă jumătate din totalul citărilor primite de publicația p în anul t (38). Documentele d'_i care citează publicările $D(p, t)$ din anul t produse de publicația analizată p , sunt numărate și sortate după data apariției $t(d'_i)$ în ordine descrescătoare. Valoarea funcției se obține scăzând din t anul apariției documentului având indicele egal cu jumătate din numărul total de documente care citează publicația p :

$$H_1(p, t) = t - t(d'_\chi), \text{ unde :}$$

$$\Leftrightarrow \chi = \lfloor |D(p, t)| / 2 \rfloor;$$

$$\Leftrightarrow D(p, t) = \{ d'_i \mid d'_i \in D, t(d'_i) \geq t(d'_{i+1}), \rho(d, d'_i) = 1, d \in D(p, t) \}$$

Funcția de înjumătățire a citărilor nu reflectă valoarea științifică a unei anumite publicații (40) ea poate da informații cu privire la politica editorială sau domeniul de cercetare – o valoare mică poate reflecta o politică editorială care pune accent pe conștientizarea actuală sau pe rapida evoluție a domeniului de cercetare, în timp ce o valoare mare poate reflecta fie accentul pe literatura de arhivă fie o evoluție lentă a domeniului (23). În practică, funcția de înjumătățire a citărilor poate fi utilizată pentru ajustarea politicilor editoriale sau intrarea pe noi segmente de cercetare.

2.12 self citedness - ponderea autocităților, în cazul în care autorii din entitatea analizată x își citează propriile publicări.

Se calculează $\rho(x, T)$ numărul total de citări pentru toate publicările entității analizate x în intervalul de timp analizat T . Se verifică de unde provin citările și se calculează numărul de citări provenind de la entitatea analizată $\rho_x(x, T)$. Se împarte al doilea număr cu primul pentru a obține ponderea autocităților:

$$c_s = \rho_x(x, T) / \rho(x, T), \text{ unde:}$$

$$\Leftrightarrow \rho_x(x, T) \equiv \rho(D(x), D(x, T)) = \sum_{d \in D(x)} \sum_{d' \in D(x, T)} \rho(d, d')$$

$$\Leftrightarrow \rho(x, T) \equiv \rho(D(x), D(T)) = \sum_{d \in D(x)} \sum_{d' \in D(T)} \rho(d, d')$$

2.13 uncitedness - ponderea publicărilor necitate ale unei entități analizate x după o anumită perioadă de timp T .

$$\text{Autocitățile sunt eliminate: } ((d, d' \in D(x, a)) \wedge (\rho(d, d') = 1)) \Rightarrow \rho(d, d') = 0$$

Se determină numărul de publicări care nu au fost citate în perioada de timp T și se împarte cu numărul total de publicări ale entității analizate din aceeași perioadă de timp:

$$p_n = |D_0(x, T)| / \Pi(x, T), \text{ unde}$$

$$\Leftrightarrow D_0(x, T) = \{ d \mid d \in D(x, T), \rho(d) = 0 \};$$

$$\Leftrightarrow |D_0(x, T)| = \sum_{d \in D(x, T)} \sum_{d' \in D} (-1)(\rho(d, d') - 1), \text{ numărul de publicări necitate ale lui } x \text{ în perioada } T;$$

$$\Leftrightarrow \Pi(x, T) = \sum_{d \in D(x, T)} \pi(d, x) = \text{numărul total de publicări ale lui } x \text{ în perioada } T.$$

Tabelul 8. Indicatori bibliometrici de performanță (1)

Definiția 19. *Funcția de impact a unui document d este o funcție $I(d, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(d, t)$ este o valoare scalară care descrie impactul tuturor referințelor la d înainte de momentul (anul) t .*

$I(d, t)$, valoarea funcției de impact a lui d la momentul t , depinde de:

- ↳ numărul $\rho(d)$ de citări ale documentului d în intervalul de timp $T = [t(d), t-1]$ unde $t(d)$ este anul publicării documentului d ;
- ↳ indicii $\alpha(d')$ și $\sigma(d, d')$ care descriu credibilitatea documentelor d' care citează pe d și respectiv relevanța acestor citări.

O variantă simplă, dar calculabilă, de definire a funcției de impact a unui document analizat $d \in D$ este:

$$I(d, t) = \sum_{d' \in D(T)} \rho(d, d'), \text{ unde:}$$

- ↳ $T = [t(d), t-1]$ este intervalul de timp analizat;
- ↳ sumarea se face pentru toate documentele d' care conțin o referință la d și au fost publicate în intervalul de timp T , $t(d') \in T$.

Definiția 20. *Funcția de impact a unei mulțimi de documente, \mathcal{D} , este o funcție $I(\mathcal{D}, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(\mathcal{D}, t)$ este o valoare scalară care descrie impactul tuturor referințelor la documentele $\vec{d} \in \mathcal{D}$ înainte de momentul (anul) t :*

$$I(\mathcal{D}, t) = \sum_{\vec{d} \in \mathcal{D}} I(\vec{d}, t), \text{ unde}$$

- ↳ $I(\vec{d}, t)$ este valoarea funcției de impact a documentului \vec{d} la momentul t ;

Definiția 21. *Funcția de impact a unui autor este o funcție $I(a, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(a, t)$ este o valoare scalară care descrie impactul tuturor documentelor publicate de autorul a înainte de momentul (anul) t :*

$$I(a, t) = \sum_{d \in D(a)} I(d, t), \text{ unde}$$

- ↳ $I(d, t)$ este valoarea funcției de impact a documentului d la momentul t ;
- ↳ $D(a)$ este mulțimea tuturor documentelor d publicate de autorul a înainte de momentul t , $t(d) \leq t-1$.

Definiția 22. *Funcția de impact a unei publicații, p , este o funcție $I(p, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(p, t)$ este o valoare scalară care descrie impactul din momentul t al tuturor documentelor publicate înainte de momentul (anul) t , în publicația p :*

$$I(p, t) = \sum_{d \in D(p)} I(d, t), \text{ unde}$$

- ↳ $I(d, t)$ este valoarea funcției de impact a documentului d la momentul t ;
- ↳ sumarea se face pentru toate documentele d publicate de publicația p înainte de momentul t , $t(d) \leq t-1$.

Definiția 23. *Funcția de impact a unui domeniu de cercetare, c , este o funcție $I(c, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I(c, t)$ este o valoare scalară care descrie impactul din momentul t al tuturor documentelor publicate înainte de momentul t în toate publicațiile din domeniul c , $p \in P(c)$.*

$$I(c, t) = \sum_{p \in P(c)} I(p, t), \text{ unde:}$$

- ↳ $I(p, t)$ este valoarea funcției de impact a publicației p la momentul t ;

↪ sumarea se face pentru toate publicațiile $p \in P(c)$.

Definiția 24. *Funcția de impact a unei entități pentru o fereastră de citare de n ani* este o funcție $I_n(x, \bullet) : N \rightarrow \mathfrak{R}_+$, unde $I_n(x, t)$ este o valoare scalară care descrie impactul din momentul t al tuturor documentelor publicate de entitatea analizată x într-un interval de timp analizat $T = [t-n, t-1]$:

$$I_n(x, t) = \sum_{d \in D(x, T)} I(d, t), \text{ unde}$$

↪ $I(d, t)$ este valoarea funcției de impact a documentului d la momentul t ;

↪ sumarea se face pentru toate documentele d publicate de entitatea x în intervalul de timp analizat, $t(d) \in T$.

Definiția 25. *Factorul de impact al unei entități analizate x pentru o fereastră de citare de n ani*, este:

$$IF_n(x) = I_n(x, t) / \Pi(x, T), \text{ unde}$$

↪ $T = [t-n, t-1]$ este intervalul de timp analizat (fereastra de citare);

↪ $I_n(x, t)$ = valoarea la momentul t a funcției de impact a entității x pentru perioada T ;

↪ $\Pi(x, T)$ = numărul total de documente publicate de entitatea x în aceeași perioadă.

2.14 Journal Impact Factor (IF) – *factorul de impact al unei publicații, pentru o fereastră de citare de 2 ani* este:

$$IF = IF_2(p) = I_2(p, t) / \Pi(p, T), \text{ unde:}$$

↪ $I_2(p, t)$ este valoarea funcției de impact a publicației p la momentul t , pentru $n = 2$;

↪ $\Pi(p, T)$ reprezintă numărul total de documente publicate de publicația p în intervalul de timp $T = [t-2, t-1]$.

2.15 5-year journal IF – *factorul de impact al unei publicații, pentru o fereastră de citare de 5 ani*, este:

$$IF_5(p) = I_5(p, t) / \Pi(p, T), \text{ unde}$$

↪ $I_5(p, t)$ este valoarea funcției de impact a publicației p la momentul t , pentru $n = 5$;

↪ $\Pi(p, T)$ reprezintă numărul total de documente publicate de publicația p în intervalul de timp analizat, $T = [t-5, t-1]$.

2.16 Immediacy Index (II) – *factorul de impact imediat* (pentru o fereastră de citare de 0 ani) al unei publicații p , măsoară importanța actuală a documentelor publicate de p :

$$II(p) = I_0(p, t) / \Pi(p, T), \text{ unde}$$

↪ $I_0(p, t)$ este valoarea funcției de impact a publicației p în momentul (anul) t , $n = 0$;

↪ $\Pi(p, T)$ reprezintă numărul total de documente publicate de p în același an $T = t$.

2.17 Journal-to-Field Impact Score (JFIS) – *factorul de impact publicație-domeniu* măsoară numărul mediu citări per document, pe o perioadă de 5 ani, pentru o anumită publicație p și compară acest număr cu cel al tuturor publicațiilor din același domeniu de cercetare c :

$$JFIS(c, p) = IF_5(p) / IF_5(c).$$

2.18 CWTS field normalized citation score (crown indicator) – indicatorul coroană, măsoară impactul științific al unui cercetător sau a unui grup de cercetare (14) și se calculează prin împărțirea sumei numerelor de citări primite de publicările d produse de entitatea analizată x în intervalul de timp analizat T în domeniul de cercetare c , la suma mediilor citărilor documentelor d' de același tip τ , din același an $t \in T$ și publicate în reviste p din același domeniu c :

$$CI(x, c) = [\sum_{d(c) \in \Pi(x, T)} \rho(d(c)) / [\sum_{d(c) \in \Pi(x, T)} (\sum_{d' \in D(\tau(d))} \rho(d')) / |D(\tau(d))|]], \text{ unde:}$$

$$\Leftrightarrow D(\tau(d)) = \{d' \in D \mid [\tau(d') = \tau(d)] \wedge [t(d') = t(d)] \wedge [c(d') = c(d)]\}, (\forall) d(c) \in \Pi(x, T).$$

$$\Leftrightarrow \tau(d) \in \{\text{review article, technical report, original research article, case report, pictorial essays}\}.$$

Tabelul 8. Indicatori bibliometrici de performanță (2)

Definiția 26. Vectorul de influență al publicațiilor [38] dintr-un set dat $P_S = \{p_s\}_{s=1 \div S}$ de publicații sursă, este vectorul propriu lider (corespuzător celei mai mari valori proprii), p^* , al matricii P unde:

$$\Leftrightarrow P = \alpha H' + (I - \alpha) v e^T \text{ este matricea asociată, conform abordării PageRank, setului } P_S.$$

$$\Leftrightarrow \text{vectorul } v \text{ este un vector cu } S \text{ componente } v_s = \Pi(s, T) / \sum_{s=1}^S \Pi(s, T) \text{ reprezentând raportul dintre numărul total de publicări produse de publicația } s \in S \text{ și numărul total de publicări produse de toate cele } S \text{ publicații sursă pentru o fereastră de citare de cinci ani, } T = [t-5, t-1]. \text{ Vectorul } v \text{ este normalizat, suma componentelor fiind } 1.$$

$$\Leftrightarrow e^T \text{ este vectorul linie } I \text{ și deci matricea } N = v e^T \text{ este o matrice cu coloane identice fiecare egală cu vectorul } v.$$

$$\Leftrightarrow \text{matricea } H' = [H']_{ij} \equiv (h'_{ij}) \text{ este obținută din matricea } H \text{ prin înlocuirea, cu vectorul } v, \text{ a coloanelor cu toate elementele nule (noduri suspendate, cazul publicațiilor care nu citează nici o altă publicație);}$$

$$\Leftrightarrow \text{matricea } H = [H]_{ij} \equiv (h_{ij}), \text{ obținută prin normalizarea matricii } Z: h_{ij} = z_{ij} / \sum_k z_{kj}$$

$$\Leftrightarrow \text{matricea } Z = [Z]_{ij} \equiv (z_{ij}) \text{ are elementele:}$$

$$z_{ij} = \begin{cases} \rho(i, j) & \text{pentru } i \neq j \\ 0 & \text{pentru } i = j \end{cases}$$

În matricea Z sunt omise autocitățile (făcând 0 toate elementele de pe diagonală).

$$\Leftrightarrow \rho(i, j) = \text{numărul de citări (de la documentele) din publicația } j \text{ la (documentele din) publicația } i, \text{ în intervalul de timp } T = [t-5, t-1]; i, j = 1 \div S$$

$$\rho(i, j) \equiv \rho(D(i, T), D(j, T)) = \sum_{d \in D(i, T)} \sum_{d' \in D(j, T)} \rho(d, d')$$

Componentele vectorului p^* sunt folosite în ponderarea valorilor de citare.

2.19 eigenfactor scor – scorul vectorului propriu [38] al publicației s , EF_s , este definit ca procent din totalul de citate ponderate pe care publicația s le primește de la cele S publicații sursă. Vectorul de scoruri Eigenfactor este:

$$EF = 100 \text{Hp}^* / \Sigma_s [\text{Hp}^*]_s$$

2.20 article influence score – scorul de influență al documentelor [38] din publicația s , AI_s , este o măsură a influenței citărilor per document ale publicației s . Scorul de influență al documentelor din publicația s este:

$$AI_s = 0,01 EF_s / v_s, \text{ unde}$$

↪ $EF_s \equiv [EF]_s$ este scorul Eigenfactor al publicației s și $v_s \equiv [v]_s$ este a s -a componentă a vectorului v .

Tabelul 8. Indicatori bibliometrici de performanță (3)

Definiția 27. *Indicele de notorietate al unei mulțimi de entități analizate, X , este o valoare scalară $\mathcal{E}(X)$ care depinde de toți sau de o parte a indicilor de notorietate $\mathcal{e}(x)$ atașați entităților din X .*

Exemplul 1. *Indicele de notorietate al unei mulțimi de documente X este un indice $\mathcal{E}(X)$ care depinde de indicii de notorietate ai editurilor și/sau publicațiilor pentru fiecare $d \in X$. În mod obișnuit $X = D(x)$ unde entitatea analizată x poate fi un autor a , un grup de cercetare g , o publicație p sau o editură e :*

$$\mathcal{E}(X) \equiv \mathcal{E}(D(x)) = \phi (\{ (\mathcal{e}(E(d)), \mathcal{e}(P(d))) \mid d \in D(x) \})$$

O variantă simplă, dar calculabilă, a definiției este:

$$\mathcal{E}(D(x)) = \Sigma_{d \in D(x)} (w_E(\mathcal{e}(E(d))) + w_P \mathcal{e}(P(d))), \text{ unde}$$

↪ $w_E + w_P = I$; $w_E, w_P \geq 0$.

Exemplul 2. *Indicele de notorietate al unui autor a este un indice $\mathcal{E}_3(a)$ care depinde de a și de afilierea acestuia, $G(a)$.*

O variantă simplă, dar evaluabilă a definiției este:

$$\mathcal{E}(a) = w_A \mathcal{e}(a) + w_G \mathcal{e}(G(a)), \text{ unde}$$

↪ $w_A + w_G = I$; $w_A, w_G \geq 0$.

Definiția 28. *Indicele de notorietate-impact al unei mulțimi de documente X , este un indicator $\mathcal{EI}(X)$ care depinde de indicele de notorietate $\mathcal{E}(X)$ și de valoarea funcției de impact $I(X, t)$, în anul de referință t .*

Cea mai simplă formă de definiție calculabilă este:

$$\mathcal{EI}(X) := w_I \mathcal{E}(X) + w_2 I(X, t), \text{ unde}$$

↪ $w_I + w_2 = I$; $w_I, w_2 \geq 0$.

3.2.3.4 Comentarii:

- ↪ Indicatorii de performanță se bazează pe ipoteza că frecvența citărilor unui articol de către alte articole reflectă calitatea acelui articol.
- ↪ Indicatorii oferă doar o imagine parțială și părtinitoare a anumitor aspecte ale vieții științifice, fără acoperirea ansamblului. Aceștia trebuie să fie completați și/sau corecțai de experții din domeniul științei și, de asemenea, interpretați dacă sunt utilizați în scopul unei evaluări sau luări de decizii.
- ↪ Se recomandă utilizarea literaturii bibliometrice pentru a identifica cât mai precis semnificația reală a indicatorilor și părtinirile lor.
- ↪ Măsurarea calității și impactului revistelor științifice, a grupurilor de cercetare sau a cercetătorilor individuali este recomandabil să fie făcută prin intermediul mai multor indicatori și nu doar prin unul singur.
- ↪ Având în vedere diferențele semnificative dintre domeniile de cercetare în ceea ce privește productivitatea, obiceiurile de citare și dinamica citărilor, se recomandă ca indicatorii bibliometrici să nu fie utilizați pentru a compara cercetători, grupuri de cercetare sau reviste din domenii diferite.
- ↪ Deoarece s-a constatat că în analiza citărilor erorile de măsurare sunt foarte importante se recomandă cu tărie multiplicarea surselor, analiza critică a rezultatelor cu implicarea experților din domeniu și, de asemenea, o atenție deosebită la ordinele de mărime al indicatorilor.
- ↪ Indicatorii numerici sunt foarte ușor manipulabili de către persoane fizice, instituții și alte părți interesate din viața științifică (cum ar fi revistele). Numărul manipulărilor crește și poate fi corelat cu efectul influenței crescânde a indicatorilor.
- ↪ Utilizarea indicatorilor bazați pe analiza citărilor nu este favorabilă asumării de riscuri științifice și inovării. O utilizare abuzivă a acestora sau, mai rău, automată ar fi un obstacol major în calea inovării.

3.3 Reconcilierea cu sursele de date

3.3.1 Sursele de date

Pentru depozitul de date al unui sistem suport pentru decizii de bibliotecă principalele surse de date pot fi:

- ↪ Bazele de date operaționale ale bibliotecii ;
- ↪ Bazele de date bibliografice, respectiv cataloagele on-line de bibliotecă ;
- ↪ Biblioteci digitale precum Europeana, WDL ș.a.
- ↪ Publicații și baze de date internaționale, on-line, pentru documentare precum Science Direct, SpringerLink, Wiley Blackwell, Taylor and Francis, DOAJ, etc.;
- ↪ Baze de date bibliometrice precum ISI Web of Science (Thomson Reuters) , Scopus (Elsevier), Google Scholar (Google), ș.a.

Reconcilierea cerințelor informaționale privind descrierile bibliografice cu sursele de date se realizează prin maparea informațiilor existente în sursele de date pe informațiile considerate necesare în modelul FRBR descris anterior. În raport cu cerințele respective

sursele de date oferă mult mai puține informații obținându-se practic un fel de surrogate ale descrierilor bibliografice după cum se poate vedea în exemplele următoare.

Exemplul 1. Descrierea bibliografică a unei *publicații* (revistă)

FRBR	Surse de date	Informația bibliografică
APw01	<i>titlu-pub</i>	Acta Electrotechnica et Informatica
APe10	<i>context-pub</i>	
APw02	<i>forma-pub</i>	J (journal)
APe02	<i>tip-pub</i>	E; P (versiune Electronica si/sau Printata)
APm14	<i>id-int-pub</i>	[EISSN: 13383957; ISSN: 13358243]
APw03	<i>data-pub</i>	2011 (start year)
APe15	<i>frecvența-pub</i>	T (trimestrial)
APm04	<i>țara</i>	SL (Slovakia)
APe04	<i>limba-pub</i>	ENG (english)
APw06	<i>domeniu-pub</i>	[Computer Science]
ASc01	<i>subiect-pub</i>	[electrical engineering, electronics and microelectronics, computer science, information technology, information systems] (keywords)
APe08	<i>volum-doc</i>	(nr. pag.)
APe09	<i>conținut-pub</i>	C (Cuprins / Full text)
APm02	<i>editor-pub</i>	[<i>id-rsp</i> (Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia)]
APm05	<i>editură-pub</i>	<i>id-edt</i> (Versita Open ; Versita Ltd, 78 York Street, London W1H 1DP, Great Britain, Tel: +44 (0) 20 7692 4902, info@versita.com)
APm03	<i>ediție-pub</i>	1
APm08	<i>colecție-pub</i>	Versita Emerging Science Publishers
APm13	<i>format-pub</i>	A4
APm15	<i>furnizor-pub</i>	Versita Open
APm16	<i>cost-pub</i>	0
APm17	<i>restricții-pub</i>	open access
APm22	<i>stare-pub</i>	A (activ)
APm23	<i>nrotare-pub</i>	Volume 11 (2011), Issue 1(mar 2011)
APm35	<i>config-pub</i>	Adobe
APm36	<i>fișier-pub</i>	pdf
APm37	<i>acces-pub</i>	on-line
APm38	<i>adrURL-pub</i>	http://versita.com/aei/

Tabelul 9. Descrierea bibliografică a unei reviste

Exemplul 2. Descrierea bibliografică a unei *publicații* (proceedings)

FRBR	Surse de date	Informația bibliografică
APw01	<i>titlu-pub</i>	Proceedings 2003 VLDB Conference
APe10	<i>context-pub</i>	29 th International Conference on Very Large Databases, Berlin, 9-12 sept, 2003
APw02	<i>forma-pub</i>	P (proceedings)
APe02	<i>tip-pub</i>	E; (electronic version)
APm14	<i>id-int-pub</i>	ISBN: 978-0-12-722442-8
APw03	<i>data-pub</i>	2003
APe15	<i>frecvența-pub</i>	
APm04	<i>țara</i>	GER (Germany)
APe04	<i>limba-pub</i>	ENG (english)
APw06	<i>domeniu-pub</i>	Databases

<i>ASc01</i>	<i>subiect-pub</i>	[very large databases] (keywords)
<i>APe08</i>	<i>volum-doc</i>	345 (pagini)
<i>APe09</i>	<i>conținut-pub</i>	[Keynotes, Ten-Year Best Paper Award, Research Sessions, Industrial Sessions, Panels, Demo Sessions, Tutorials] (capitole)
<i>APm02</i>	<i>editor-pub</i>	Johann-Christoph Freytag, Peter Lockemann, Serge Abiteboul, Michael Carey, Patricia Selinger; Andreas Heuer (editors)
<i>APm05</i>	<i>editură-pub</i>	Elsevier Inc (publisher)
<i>APm03</i>	<i>ediție-pub</i>	I
<i>APm08</i>	<i>colecție-pub</i>	
<i>APm13</i>	<i>format-pub</i>	A4
<i>APm15</i>	<i>furnizor-pub</i>	
<i>APm16</i>	<i>cost-pub</i>	
<i>APm17</i>	<i>restricții-pub</i>	organization's agreement with publisher
<i>APm22</i>	<i>stare-pub</i>	A (activ)
<i>APm23</i>	<i>notare-pub</i>	
<i>APm35</i>	<i>config-pub</i>	Adobe
<i>APm36</i>	<i>fișier-pub</i>	PDF
<i>APm37</i>	<i>acces-pub</i>	on-line
<i>APm38</i>	<i>adrURL-pub</i>	http://www.sciencedirect.com/science/book/9780127224428

Tabelul 10. Descrierea bibliografică a unui *proceedings*.**Exemplul 3.** Descrierea bibliografică a unei *publicații* (monografie)

FRBR	Surse de date	Informația bibliografică
<i>APw01</i>	<i>titlu-pub</i>	Handbook on Decision Support Systems V1: Basic Themes V2: Variations
<i>APe10</i>	<i>context-pub</i>	
<i>APw02</i>	<i>forma-pub</i>	M (monografie)
<i>APe02</i>	<i>tip-pub</i>	E; P (electronic version: eBook, print version: Hardcover)
<i>APm14</i>	<i>id-int-pub</i>	ISBN: 978-3-540-48712-8; eISBN: 978-3-540-48713-5
<i>APw03</i>	<i>data-pub</i>	2008
<i>APe15</i>	<i>frecvența-pub</i>	
<i>APm04</i>	<i>țara</i>	GER (Germany)
<i>APe04</i>	<i>limba-pub</i>	ENG (english)
<i>APw06</i>	<i>domeniu-pub</i>	[Operations Research ; Decision Theory]
<i>ASc01</i>	<i>subiect-pub</i>	[Decision Support Systems] (keywords)
<i>APe08</i>	<i>volum-doc</i>	854; 800 (pagini)
<i>APe09</i>	<i>conținut-pub</i>	[V1: Foundations of DSS, 1-8; DSS Fundamentals, 9-16; Multiparticipant DSS, 17-23; Intelligent DSS, 24-31; Effects of computer based DS, 32-36] [V2: Time and space issue for DS, 37- 42; Scope of DS, 43- 48; Developing and managing DSS, 49- 53; DS cases and applications, 54- 63; DS horizons: 64-71] (parts and chapters)
<i>APm02</i>	<i>editor-pub</i>	[Burstein, Frada; Holsapple, Clyde W.] (editors) Professor Frada Burstein, Center for Organizational and Social Informatics, Faculty of Information Technology, Monash University P.O. Box 197 Caulfield East, 3145, Victoria, Australia, Frada.Burstein@infotech.monash.edu.au Professor Clyde W. Holsapple, Gatton College of Business and Economics, University of Kentucky, 425B, Gatton Building, Lexington KY 40506-0034, USA, cwhols@uky.edu
<i>APm05</i>	<i>editură-pub</i>	Springer-Verlag Berlin Heidelberg (publisher)
<i>APm03</i>	<i>ediție-pub</i>	I
<i>APm08</i>	<i>colecție-pub</i>	International Handbooks on Information Systems (Series)
<i>APm13</i>	<i>format-pub</i>	A4
<i>APm15</i>	<i>furnizor-pub</i>	Springer
<i>APm16</i>	<i>cost-pub</i>	V1 eBook 189,99 € (gross) price ISBN 978-3-540-48713-5 PDF Hardcover 106,95 € (gross) price ISBN 978-3-540-48712-8
<i>APm17</i>	<i>restricții-pub</i>	organization's agreement with publisher

<i>APm22</i>	<i>stare-pub</i>	
<i>APm23</i>	<i>notare-pub</i>	
<i>APm35</i>	<i>config-pub</i>	Adobe
<i>APm36</i>	<i>fișier-pub</i>	PDF
<i>Apm37</i>	<i>acces-pub</i>	on-line
<i>APm38</i>	<i>adrURL-pub</i>	http://www.springer.com/978-3-540-48712-8

Tabelul 11. Descrierea bibliografică a unei monografii

Exemplul 4. Descrierea bibliografică a unei publicații (monografie, OPAC)

FRBR	Surse de date	Informația bibliografică
<i>Apm37</i>	Cota	II 924314
<i>Apm02</i>	Autor	Filip, Florin Gheorghe
<i>Apw01</i>	Titlul	Sisteme suport pentru decizii
<i>Apm02</i>	Respons. primară	acad. Florin Gheorghe Filip
<i>Apm04</i>	Localitate	București
<i>Apm05</i>	Editură	Editura Tehnică
<i>Apw03</i>	An	2007
<i>APe08</i>	Descriere fizică	364 p. : il. ; 24 cm
<i>Apm08</i>	Serie	Tehnologia informației
<i>APe04</i>	Limba	rum
<i>Apm04</i>	Țara de publicare	RO
<i>APe09</i>	Bibliografie	p. 320-346
<i>Apm14</i>	ISBN/Pret	978-973-31-2308-8
<i>Apm08</i>	Titlul seriei	Tehnologia informației
<i>APe09</i>	Subiect termen nec.	Tehnologia informației
		Sistem informațional
		Sistem suport
		Informatică
		Management

Tabelul 12. Descrierea bibliografică a unei monografii, format UNIMARC

Exemplul 5. Descrierea bibliografică a unui document (articol)

FRBR	Surse de date	Informația bibliografică
<i>APw01</i>	<i>titlu-doc</i>	A fuzzy incremental clustering approach to hybrid data discovery
<i>APw02</i>	<i>forma-doc</i>	A (Articol / Capitol)
<i>APe02</i>	<i>tip-doc</i>	E (versiune Electronică / Tiparita)
<i>APw03</i>	<i>data-doc</i>	23 – 10 – 2012 (data publicare online)
<i>APe04</i>	<i>limba-doc</i>	ENG (english)
<i>APw06</i>	<i>domeniu-doc</i>	[data mining]
<i>APe05</i>	<i>subiect-doc</i>	[incremental clustering; fuzzy; agents; hybrid data] (keywords)
<i>APe08</i>	<i>volum-doc</i>	8 (pagini)
<i>APe09</i>	<i>rezumat-doc</i>	A (abstract)
<i>APi06</i>	<i>stare-doc</i>	T (full text)
<i>APm02</i>	<i>resp-doc</i>	[<i>id-aut</i> (Găceanu, Radu D.) ; <i>id-aut</i> (Pop, Horia F.)] ; [<i>id-afl</i> (Computer Science Department, Babes-Bolyai University, Cluj-Napoca, Romania)]
<i>APm03</i>	<i>ed-pub</i>	[<i>id-pub</i> (Acta Electrotechnica et Informatica); Vol. 12, No. 2, Oct 2012, 16–23]
<i>APm13</i>	<i>format-doc</i>	A4
<i>APm16</i>	<i>cost-doc</i>	null
<i>APm17</i>	<i>restricții-doc</i>	free

<i>APm35</i>	<i>config-doc</i>	Adobe
<i>APm36</i>	<i>fișier-doc</i>	PDF; 190 K
<i>APm37</i>	<i>acces-doc</i>	on-line
<i>APm38</i>	<i>adrURL-doc</i>	http://www.degruyter.com/view/j/aeeci.2012.12.issue-2/v10198-012-0010-x/v10198-012-0010-x.xml

Tabelul 13. Descrierea bibliografică a unui articol de revistă

Exemplul 6. Descrierea bibliografică a unui *document* (capitol de monografie)

FRBR	Surse de date	Informația bibliografică
<i>APw01</i>	<i>titlu-doc</i>	DSS Architecture and Types
<i>APw02</i>	<i>forma-doc</i>	C (Articol / Capitol)
<i>APe02</i>	<i>tip-doc</i>	E (versiune Electronică sau Printata)
<i>APw03</i>	<i>data-doc</i>	2008 (publicare)
<i>APe04</i>	<i>limba-doc</i>	ENG (english)
<i>APw06</i>	<i>domeniu-doc</i>	[DSS]
<i>APe05</i>	<i>subiect-doc</i>	[Architecture; Decision support system; DSS; Framework; Knowledge system; Language system; Presentation system; Problem-processing system] (keywords)
<i>APe08</i>	<i>volum-doc</i>	27 (pagini)
<i>APe09</i>	<i>rezumat-doc</i>	A (abstract)
<i>APi06</i>	<i>stare-doc</i>	F (full text)
<i>APm02</i>	<i>resp-doc</i>	[<i>id-aut</i> (Clyde W. Holsapple)] [<i>id-afl</i> (School of Management, Gatton College of Business and Economics, University of Kentucky, Lexington, KY, USA)]
<i>APm03</i>	<i>ed-pub</i>	<i>id-pub</i> (Handbook on Decision Support Systems Vol 1: Basic Themes, Part II: DSS Fundamentals, Chapter 9, pp 163- 189)
<i>APm13</i>	<i>format-doc</i>	A4
<i>APm16</i>	<i>cost-doc</i>	25 (€)
<i>APm17</i>	<i>restricții-doc</i>	none
<i>APm35</i>	<i>config-doc</i>	Adobe
<i>APm36</i>	<i>fișier-doc</i>	PDF; 190 K
<i>APm37</i>	<i>acces-doc</i>	on-line
<i>APm38</i>	<i>adrURL-doc</i>	http://www.degruyter.com/view/j/aeeci.2012.12.issue-2/v10198-012-0010-x/v10198-012-0010-x.xml

Tabelul 14. Descrierea bibliografică a unui capitol dintr-o monografie

3.3.2 Surogat bibliografic documente

	FRBR	Surse de date		FRBR	Surse de date		FRBR	Surse de date
1	<i>APw01</i>	<i>titlu-doc</i>	8	<i>APe08</i>	<i>volum-doc</i>	15	<i>APm17</i>	<i>restricții-doc</i>
2	<i>APw02</i>	<i>forma-doc</i>	9	<i>APe09</i>	<i>rezumat-doc</i>	16	<i>APm35</i>	<i>config-doc</i>
3	<i>APe02</i>	<i>tip-doc</i>	10	<i>APi06</i>	<i>stare-doc</i>	17	<i>APm36</i>	<i>fișier-doc</i>
4	<i>APw03</i>	<i>data-doc</i>	11	<i>APm02</i>	<i>resp-doc</i>	18	<i>APm37</i>	<i>acces-doc</i>
5	<i>APe04</i>	<i>limba-doc</i>	12	<i>APm03</i>	<i>ed-pub</i>	19	<i>APm38</i>	<i>adrURL-doc</i>
6	<i>APw06</i>	<i>domeniu-doc</i>	13	<i>APm13</i>	<i>format-doc</i>			
7	<i>APe05</i>	<i>subiect-doc</i>	14	<i>APm16</i>	<i>cost-doc</i>			

Tabelul 15. Maparea pe sursele de date a descrierii bibliografice - document

3.3.3 Surogat bibliografic publicații

	FRBR	Surse de date		FRBR	Surse de date		FRBR	Surse de date
1	<i>APw01</i>	<i>titlu-pub</i>	10	<i>APw06</i>	<i>domeniu-pub</i>	19	<i>APm15</i>	<i>furnizor-pub</i>
2	<i>APe10</i>	<i>context-pub</i>	11	<i>ASc01</i>	<i>subiect-pub</i>	20	<i>APm16</i>	<i>cost-pub</i>
3	<i>APw02</i>	<i>forma-pub</i>	12	<i>APe08</i>	<i>volum-doc</i>	21	<i>APm17</i>	<i>restricții-pub</i>
4	<i>APe02</i>	<i>tip-pub</i>	13	<i>APe09</i>	<i>rezumat-pub</i>	22	<i>APm22</i>	<i>stare-pub</i>
5	<i>APm14</i>	<i>id-int-pub</i>	14	<i>APm02</i>	<i>editor-pub</i>	23	<i>APm23</i>	<i>nrotare-pub</i>
6	<i>APw03</i>	<i>data-pub</i>	15	<i>APm05</i>	<i>editură-pub</i>	24	<i>APm35</i>	<i>config-pub</i>
7	<i>APe15</i>	<i>frecvența</i>	16	<i>APm03</i>	<i>ediție-pub</i>	25	<i>APm36</i>	<i>fișier-pub</i>
8	<i>APm04</i>	<i>țara</i>	17	<i>APm08</i>	<i>colecție-pub</i>	26	<i>APm37</i>	<i>acces-pub</i>
9	<i>APe04</i>	<i>limba-pub</i>	18	<i>APm13</i>	<i>format-pub</i>	27	<i>APm38</i>	<i>adrURL-pub</i>

Tabelul 16. Maparea pe sursele de date a descrierii bibliografice - publicație

3.4 Depozitarea datelor

3.4.1 Identificare fapte

Pentru mediul decizional al unei biblioteci subiectele majore de interes sunt:

- ↪ Serviciile de bibliotecă;
- ↪ Aparițiile editoriale;
- ↪ Calitatea publicațiilor;

3.4.2 Definiere dimensiuni

Perspectivile de analiză necesare pentru faptele identificate sunt următoarele:

- ↪ Pentru serviciile de bibliotecă:
 - ↳ timpul;
 - ↳ operațiile;
 - ↳ utilizatorii.
- ↪ Pentru aparițiile editoriale:
 - ↳ timpul;
 - ↳ publicările;
 - ↳ autorii;
 - ↳ editorii;
 - ↳ publicațiile;
 - ↳ subiectele.
- ↪ Pentru calitatea publicațiilor:
 - ↳ timpul;
 - ↳ publicările;
 - ↳ autorii;
 - ↳ referințele;
 - ↳ publicațiile;
 - ↳ subiectele.

3.4.3 Definiere ierarhii dimensionale:

Dimensiuni $D \in \mathcal{D}$	Nume de niveluri $\lambda \in \Lambda$	Căi de agregare $\rho \subseteq \Lambda \times \Lambda$	Descrieri de niveluri $\delta \in \Delta_\lambda$
Timp	perioadă	●	-◇ fereastra de analiză
	an	● ↑	-◇ anul
	semestru	● ↑	-◇ semestrul
	trimestru	● ↑	-◇ trimestrul
	lună	● ↑	-◇ luna
	zi	● ↑	-◇ data
Document	format	● ↑	-◇ format-document
	tip	● ↑	-◇ tip-document
	document	● ↑	-◇ titlu -◇ limba
Autor	țară	● ↑	-◇ nume țară
	localitate	● ↑	-◇ nume oraș
	afiliere autor	● ↑	-◇ nume instituție -◇ adresă
	autor	● ↑	-◇ nume autor -◇ profesie -◇ adresă
Editor	țară	● ↑	-◇ nume țară
	localitate	● ↑	-◇ nume oraș
	afiliere editor	● ↑	-◇ nume instituție -◇ adresă
	editor	● ↑	-◇ nume editor -◇ profesie -◇ adresă
Publicație	țara	● ↑	-◇ nume țară
	oraș	● ↑	-◇ nume localitate
	editură	● ↑	-◇ nume editură -◇ adresă
	publicație	● ↑	-◇ titlu -◇ limba -◇ frecvența
















<i>Subiect</i>	<i>domeniu</i>		-◇ <i>nume de domeniu</i>
	<i>subdomeniu</i>		-◇ <i>nume de subdomeniu</i> -◇ <i>listă de descriptori</i>
	<i>subiect</i>		-◇ <i>nume de subiect</i> -◇ <i>listă de descriptori</i>
<i>Operație</i>	<i>sistem</i>		-◇ <i>nume instituție</i>
	<i>proces</i>		-◇ <i>nume proces</i>
	<i>activitate</i>		-◇ <i>nume activitate</i>
	<i>compartiment</i>		-◇ <i>nume compartiment</i>
	<i>post</i>		-◇ <i>nume angajat</i> -◇ <i>funcție angajat</i>
	<i>operație</i>		-◇ <i>nume operație</i> -◇ <i>cod operație</i>
<i>Utilizator</i>	<i>continuitate</i>		-◇ <i>re-înscris/nou-înscris</i>
	<i>naționalitate</i>		-◇ <i>română/altele</i>
	<i>gen</i>		-◇ <i>masculin/feminin</i>
	<i>vârstă</i>		-◇ <i>nume categorie-vârstă</i>
	<i>ocupație</i>		-◇ <i>nume statut-ocupațional</i>
	<i>utilizator</i>		-◇ <i>cod</i> -◇ <i>nume/număr permis</i>

Figura 3. Dimensiuni, niveluri dimensionale și căi de agregare

3.4.4 Definiere măsuri

Aspectele specifice și măsurabile ale faptelor, relevante pentru analiză, la nivelul minim de granularitate, sunt:

- ↳ Pentru serviciile de bibliotecă: indicii de selecție (s) și de realizare (r); duratele (τ) și costurile (c) unitare ale operațiilor;
- ↳ Pentru aparițiile editoriale: indicii de publicare (π) și de cotare (ψ);
- ↳ Pentru calitatea publicațiilor: indicii de notorietate (ε), de citare (ρ) și, eventual, de relevanță a citărilor (σ).

3.4.5 Setul de interogări preliminare

Sistemele de indicatori (operaționali, de performanță și bibliometrici) construite în capitoul anterior reprezintă de fapt cerințe ale utilizatorilor și constituie setul de interogări preliminare (*preliminary workload*) la care trebuie să poată răspunde depozitul de date.

3.4.6 Cuburi de date

Schemele cuburilor de date sunt reprezentate prin diagrame specifice în care: faptele sunt reprezentate prin dreptunghiuri; dimensiunile sunt reprezentate prin dreptunghiuri rotunjite; măsurile sunt reprezentate prin cercuri.

Cuburile de date cu dimensiunile din figura 3 corespunzătoare faptelor analizate (servicii, publicări și citări) sunt reprezentate în continuare:

↪ Schema „cubului de date” **Servicii de bibliotecă**:

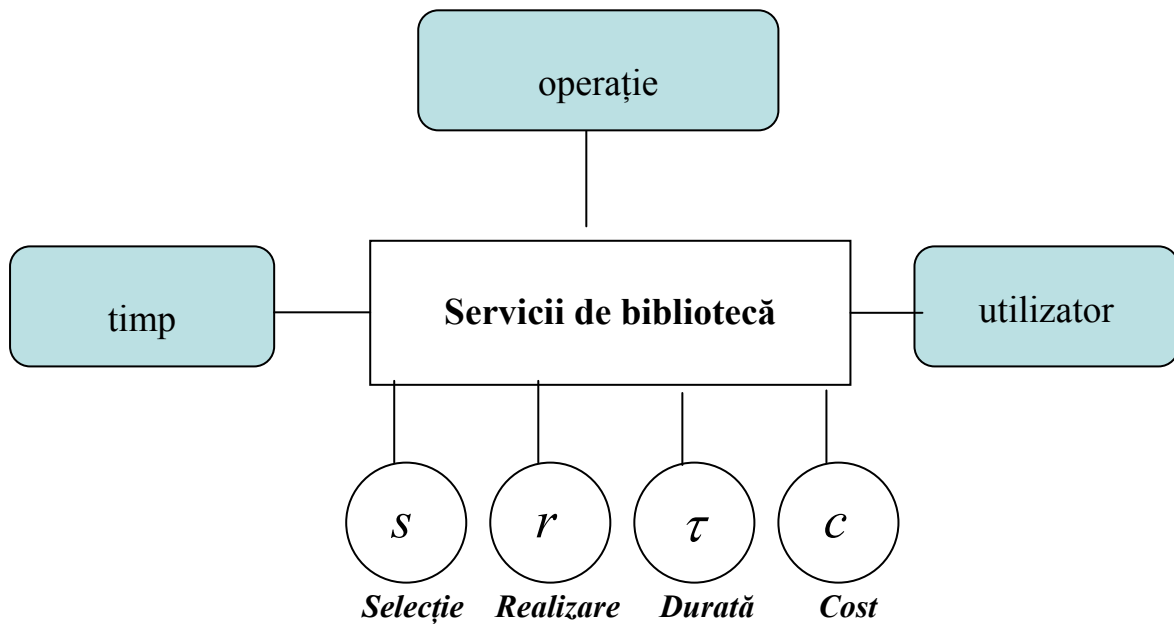


Figura 4. Cub de date privind serviciile bibliotecare

↪ Schema „cubului de date” **Publicări**:

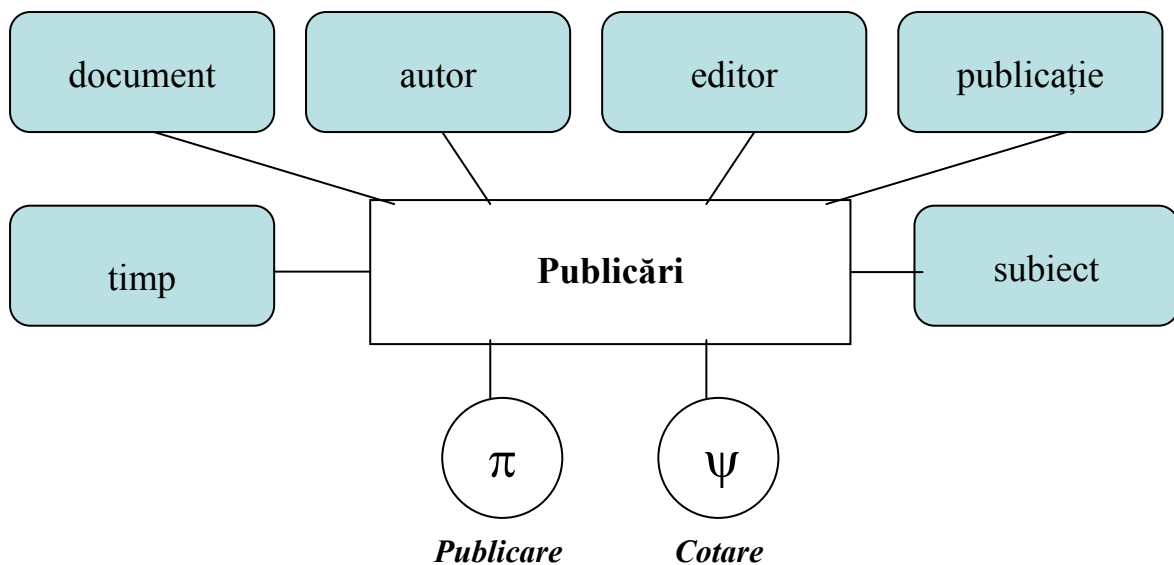


Figura 5. Cub de date privind aparițiile editoriale

↳ Schema „cubului de date” **Citări**:

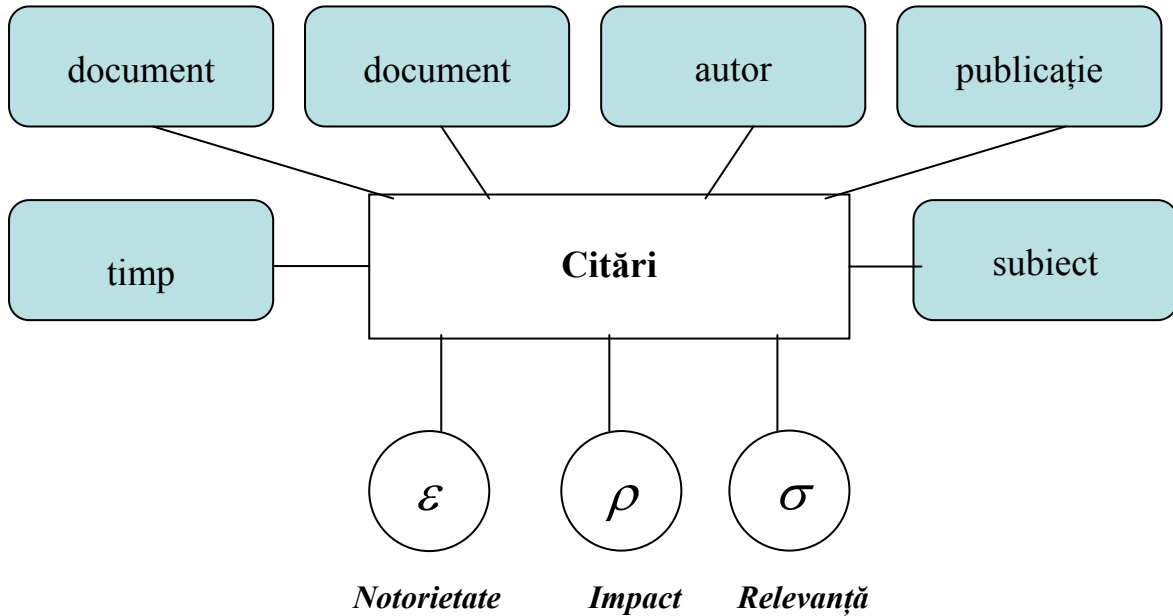


Figura 6. Cub de date privind calitatea publicărilor

3.4.7 Schema conceptuală a depozitului de date

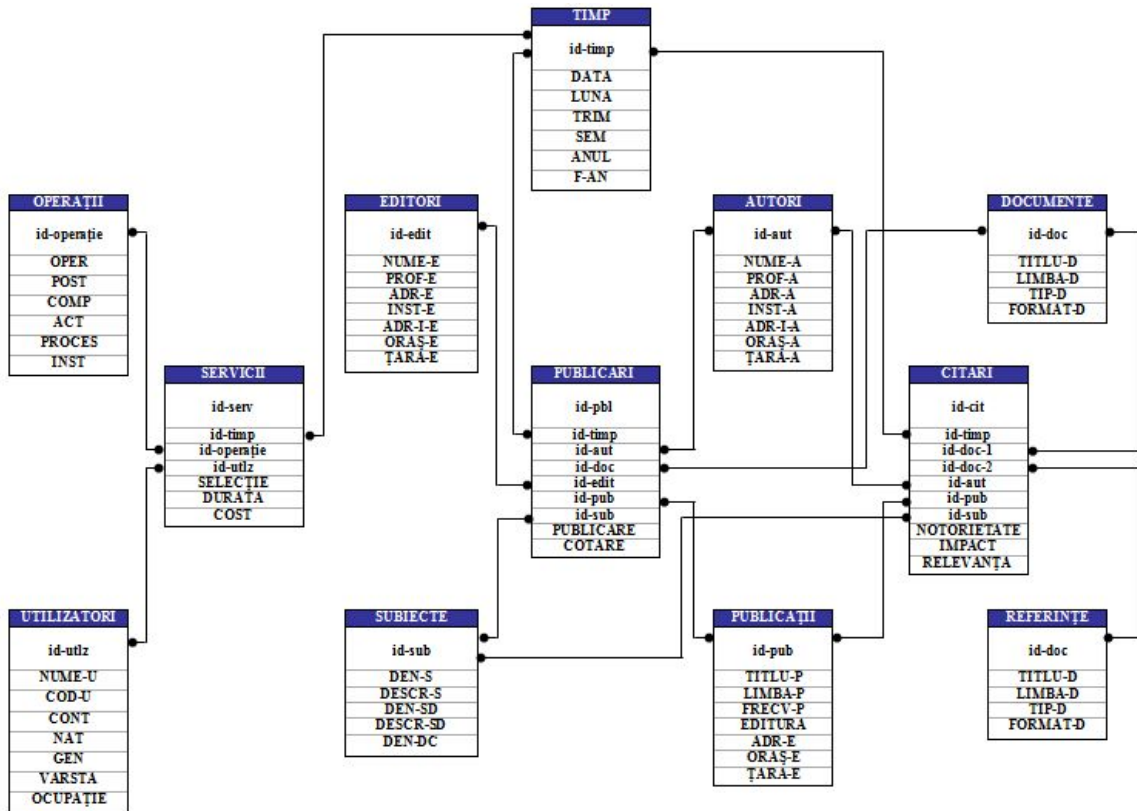


Figura 7. Schema „constelație” a depozitului de date

3.5 Descoperirea cunoștințelor din date

3.5.1 Nucleu de autoritate

Definirea de nuclee de autoritate în diferite mulțimi de entități implică definirea unei proceduri generale, cât mai simplă posibil, pentru a genera ierarhii bibliometrice. Ideea fundamentală a abordării este de a considera entitățile analizate ca părți ale unui sistem, fiecare entitate putând fi caracterizată de către celelalte. Elementul de legătură este mulțimea de documente. Compararea a două entități înseamnă să se selecteze din mulțimea de documente o secțiune corespunzătoare și să se efectueze comparația în interiorul acelei secțiuni.

În continuare este schițată o procedură pentru definirea unei ierarhizări a autorilor care au tratat un anumit subiect, definirea de proceduri similare pentru alte entități precum publicațiile, editurile sau organizațiile științifice fiind simple exerciții.

Un dezavantaj al abordării ar putea fi numărul mare de ponderi, dar procedura poate funcționa, de asemenea, cu ponderi elementare, booleene.

Fie s un anumit subiect de interes și fie \mathcal{S} o mulțime de subiecte care conține pe s . Mulțimea \mathcal{S} conține subiectul s dar include și alte subiecte apropiate din același (sub)domeniu pentru a evita căutarea într-un set prea îngust de subiecte.

Se dorește o ierarhizare în interiorul mulțimii de autori în subiectul s .

Pentru punerea în aplicare a procedurii se presupune că mulțimea documentelor care abordează subiectul s nu este vidă, $D(s) \neq \emptyset$ și, de asemenea, că există informațiile necesare, listele de clasificare notorietate-expert și mulțimile de ponderi.

Pasul-1. Inițializarea procedurii

- ↪ s := subiectul de interes
- ↪ \mathcal{S} := mulțimea de subiecte ce conține pe s
- ↪ \mathcal{A}^+ := listele autorizate cu autorii de notorietate
- ↪ \mathcal{W} := mulțimea de valori pentru ponderi

Pasul-2. Selecția documentelor care tratează subiectul s

- ↪ $D(s) := \{d(s)\}$

Pasul-3. Selecția documentelor care citează documente din $D(s)$

- ↪ $R(D(s)) := \{d' \mid \rho(d, d') = 1, d \in D(s)\}$

Pasul-4. Generarea secțiunii de documente privind subiectul s

- ↪ $\mathcal{D} := D(s) \cup R(D(s))$

Pasul-5. Generarea mulțimii de autori

- ↪ $\mathcal{A} := A(\mathcal{D}) \cup \mathcal{A}^+$, unde:
 - ↳ $A(\mathcal{D}) = \{a \mid a \in A(d); d \in \mathcal{D}\}$
 - ↳ $\mathcal{A}^+ = \{a \mid \alpha(a) > 0\}$

Pasul-6. Determinarea indicilor de notorietate □ pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathcal{A}

↳ $\mathcal{N}(\mathcal{S}, \mathcal{A}) := \{ \mathcal{E}(D(\mathcal{S}, a)) \mid a \in \mathcal{A} \}$, unde:

$$\hookrightarrow D(\mathcal{S}, a) = D(\mathcal{S}) \cap D(a)$$

$$\hookrightarrow \mathcal{E}(D(\mathcal{S}, a)) = \sum_{d \in D(\mathcal{S}, a)} (w_E \mathcal{E}(E(d)) + w_P \mathcal{E}(P(d)))$$

Pasul-7. Determinarea funcțiilor de impact / pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathcal{A}

↳ $\mathcal{I}(\mathcal{S}, \mathcal{A}) := \{ I(D(\mathcal{S}, a), t) \mid a \in \mathcal{A} \}$, unde:

$$\hookrightarrow D(\mathcal{S}, a) = D(\mathcal{S}) \cap D(a)$$

$$\hookrightarrow I(D(\mathcal{S}, a), t) = \sum_{d \in D(\mathcal{S}, a)} I(d, t)$$

Pasul-8. Determinarea indicilor de notorietate-impact \mathcal{E} pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathcal{A}

↳ $\mathcal{Y}(\mathcal{S}, \mathcal{A}) := \{ \mathcal{E}(D(\mathcal{S}, a)) \mid a \in \mathcal{A} \}$, unde:

$$\hookrightarrow \mathcal{E}(D(\mathcal{S}, a)) = w_1 \mathcal{E}(D(\mathcal{S}, a)) + w_2 I(D(\mathcal{S}, a), t)$$

Pasul-9. Furnizarea, în ordine descrescătoare, a valorilor din $\mathcal{Y}(\mathcal{S}, \mathcal{A})$

↳ Sort descending ($\mathcal{Y}(\mathcal{S}, \mathcal{A})$)

Pasul-10. Finalizare

↳ Delimitare nucleu de autoritate;

↳ Interpretări;

↳ Alte prelucrări (deterministe și/sau probabiliste).

Figura 8. Algoritm de generare nucleu de autoritate

3.5.2 Gruparea bazată pe densitate

Gruparea constă în identificarea de grupuri sau clustere într-o mulțime de date. Clusterizarea bazată pe densitate încearcă să identifice și să separe regiunile dense (foarte populate) ale unei mulțimi de puncte, \mathcal{P} , dintr-un spațiu multidimensional (*).

Densitatea este definită ca fiind numărul de puncte existente într-o regiune de căutare, de dimensiune *eps* specificată, numită *eps-vecinătate*.

Un punct este considerat *punct de bază* dacă *eps-vecinătatea* sa conține mai multe puncte decât un număr, *MinPts*, specificat. Punctele de bază sunt în interiorul unui cluster.

Un punct este considerat *punct de frontieră* dacă *eps-vecinătatea* sa conține un număr de puncte mai mic decât *MinPts* dar se află în *eps-vecinătatea* unui punct de bază.

Un punct este considerat *punct de zgomot* dacă nu este nici punct de bază și nici punct de frontieră.

(*) KRIEGEL, H.-P.; KRÖGER, P.; SANDER, J.; ZIMEK, A. (2011) *Density-based clustering*. In WIREs Data Mining and Knowledge Discovery, **1** (3), pp 231–240

Definirea clusterelor se bazează pe noțiunea de accesibilitate în densitate.

Un punct Q este *direct accesibil* în densitate dintr-un alt punct P dacă punctul Q este conținut în *eps-vecinătatea* lui P și dacă P este punct de bază. P și Q fac parte din același cluster.

Un punct Q este *accesibil* în densitate dintr-un alt punct P dacă există o secvență de puncte P_1, \dots, P_n cu $P_1 = P$ și $P_n = Q$ în care fiecare punct P_{i+1} este direct accesibil în densitate din punctul P_i .

Relația de accesibilitate în densitate nu este simetrică. Dacă punctul Q ar fi situat la frontiera unui cluster, având deci insuficient de multe puncte vecine pentru a conta ca punct de bază, procesul de găsimă a unui drum între Q și P s-ar întrerupe acesta oprindu-se la primul punct care nu este punct de bază. Începând însă procesul din punctul P drumul ar ajunge la Q și procesul s-ar opri aici, acesta fiind primul punct găsit care nu este punct de bază. Datorită acestei asimetrii, a fost necesară introducerea noțiunii de *conectare* în densitate.

Două puncte P și Q sunt *conectate* în densitate dacă există un punct O astfel încât ambele puncte P și Q sunt accesibile în densitate din O . Conectarea în densitate este simetrică.

Un *cluster* este o submulțime de puncte a lui \mathcal{P} care satisface două proprietăți:

- ↳ Toate punctele din cluster sunt reciproc conectate în densitate.
- ↳ Dacă un punct este conectat în densitate cu orice punct de cluster atunci aceasta aparține clusterului.

Algoritm de clusterizare, propus de Martin Ester, Hans-Peter Kriegel, Jörg Sander și Xiaowei Xu în 1996, [40] începe cu un punct de pornire arbitrar, care nu a fost vizitat. Sunt căutate punctele din *eps-vecinătatea* sa și în cazul în care punctul este un punct de bază un nou cluster este pornit. În caz contrar, punctul este etichetat ca zgomot. Acest punct ar putea fi ulterior regăsit într-o *eps-vecinătate* a unui alt punct de bază și atunci este inclus în acel cluster. Dacă un punct s-a dovedit a fi un punct de bază al unui cluster atunci întreaga sa *eps-vecinătate* este, de asemenea, parte din acel cluster. Sunt incluse în cluster toate punctele care se găsesc în interiorul *eps-vecinătății* precum și punctele din *eps-vecinătățile* acestora atunci când acestea sunt, de asemenea, puncte de bază. Acest proces continuă până când clusterul de conectare în densitate este complet descoperit. În continuare, un alt punct nevizitat este procesat, ceea ce duce la descoperirea unui alt cluster sau zgomot ș.a.m.d.

Parametrii *eps* și *MinPts* reprezintă o estimare a densității punctelor din cluster și trebuie să fie specificați de către utilizator.

Dacă *eps* este ales prea mic, o mare parte din puncte nu vor fi grupate în timp ce, pentru o valoare prea mare, grupurile vor fuziona și majoritatea punctelor vor fi în același cluster.

Valoarea $MinPts = 1$ nu are sens, fiecare punct va fi un grup. Valoarea $MinPts = 2$ va da același rezultat cu clasificarea ierarhică dar algoritmul bazat pe densitate este mult mai rapid. Valorile mai mari sunt mai potrivite pentru mulțimi de date cu zgomot și vor produce mai multe cluster semnificative. De regulă *MinPts* poate fi derivat din numărul de dimensiuni d ale setului de date: $MinPts \geq d + 1$.

Descrierea algoritmului (pseudocod) :

Pasul-1. Inițializare

- ↪ eps := distanța aleasă pentru eps -vecinătate (regiuneadeCăutare)
- ↪ $MinPts$:= numărul minim de puncte ales pentru a forma un cluster
- ↪ \mathcal{M} := matricea clusterelor

Pasul-2. Prelucrări preliminare

- ↪ se selectează elementele mulțimii \mathcal{P}
- ↪ se determină coordonatele punctelor $P \in \mathcal{P}$ în spațiul multidimensional
- ↪ se determină distanțele dintre puncte
- ↪ se constituie regiunile de căutare

Pasul-3. Clusterizare(\mathcal{P} , eps , $MinPts$)

- ↪ $C = 0$
pentru fiecare punct „nevizitat” P din mulțimea \mathcal{P}
se marchează P ca „vizitat”
regiuneadeCăutare(P , eps , $NrPts(P)$, $Pts(P)$)
dacă $NrPts(P) < MinPts$
expandareZgomote($Pts(P)$, eps , $MinPts$, 0)
altfel
 $C = C + 1$ (clusterul următor)
expandareCluster(P , eps , $MinPts$, C)
- ↪ **expandareCluster**(P , eps , $MinPts$, C)
se adaugă punctul P la clusterul C ($P \in C$)
pentru fiecare punct Q din mulțimea \mathcal{P}
dacă $Q \in C$ și Q nu este „membru” al clusterului C
se marchează Q ca „membru” al clusterului C
regiuneadeCăutare(Q , eps , $NrPts(Q)$, $Pts(Q)$)
clusterul $C =$ clusterul C reunit cu $Pts(Q)$
se marchează punctele din $Pts(Q)$ ca „vizitate”
- ↪ **expandareZgomote**($Pts(P)$, eps , $MinPts$, 0)
pentru fiecare punct Q din $Pts(P)$
regiuneadeCăutare(Q , eps , $NrPts(Q)$, $Pts(Q)$)
dacă $(\forall)Q \in Pts(P)$, $NrPts(Q) < MinPts$
clusterul $0 =$ clusterul 0 reunit cu $Pts(Q)$
se marchează punctele din $Pts(P)$ ca „zgomote” (membri al „clusterului” 0)

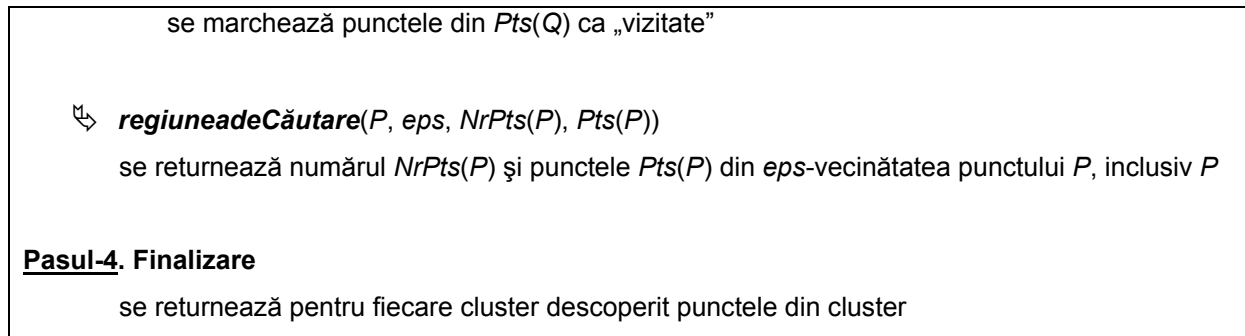


Figura 9. Algoritm de clusterizare bazată pe densitate

3.5.3 Programul CBD

Pentru algoritmul de Clusterizare Bazată pe Densitate a fost realizat un program **CBD** implementat sub sistemul MS-ACCESS. Programul **CBD**:

- ↪ nu are nevoie să i se specifice à priori numărul de grupuri;
- ↪ necesită doar doi parametri și este insensibil la ordonarea punctelor;
- ↪ poate descoperi grupuri de formă de arbitrară;
- ↪ poate descoperi un grup complet înconjurat de un alt grup;
- ↪ poate separa zgomotele.

În continuare sunt prezentate două exemple: unul ilustrativ și celălalt bazat pe un set de date preluat din [8] (Performance Indicators of General Radiology Journals, Sorted by Decreasing IF).

Exemplul 1. $\mathcal{P}_I \subset \mathbb{R}^2$ este o mulțime de puncte din plan.

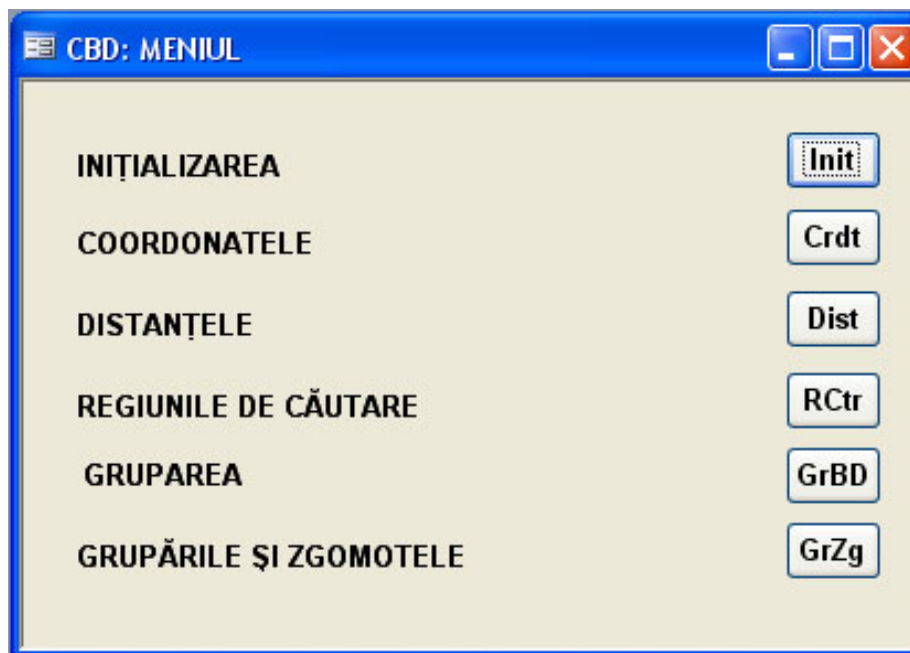


Figura 10. Meniul programului CBD

Valorile alese pentru cei doi parametri ai programului sunt: $eps = 2,237$; $MinPts = 3$

$$\mathcal{P}_1 = \{ P_1(4, 9), P_2(5, 11), P_3(5, 14), P_4(6, 8), P_5(7, 4), P_6(7, 6), P_7(7, 17), P_8(8, 8), P_9(8, 14), P_{10}(9, 11), P_{11}(9, 16), P_{12}(9, 18), P_{13}(10, 9), P_{14}(10, 14), P_{15}(11, 17), P_{16}(13, 10), P_{17}(13, 12), P_{18}(14, 15), P_{19}(15, 17), P_{20}(16, 14), P_{21}(17, 16), P_{22}(18, 10), P_{23}(18, 12) \}$$

Coordonate	c1	c2
P1	4	9
P2	5	11
P3	5	14
P4	6	8
P5	7	4
P6	7	6
P7	7	17
P8	8	8
P9	8	14
P10	9	11
P11	9	16
P12	9	18
P13	10	9
P14	10	14
P15	11	17
P16	13	10
P17	13	12
P18	14	15
P19	15	17
P20	16	14
P21	17	16
P22	18	10
P23	18	12

Figura 11. Coordonatele punctelor norului $\mathcal{P}_1 \subset \mathbb{R}^2$

CBD: DISTANȚELE																							
*	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23
P1		2,24	5,1	2,24	5,83	4,24	8,54	4,12	6,40	5,39	8,60	10,3	6	7,81	10,6	9,06	9,49	11,7	13,6	13	14,8	14,0	14,3
P2	2,236		3	3,16	7,28	5,39	6,32	4,24	4,24	4	6,40	8,06	5,39	5,83	8,49	8,06	8,06	9,95	11,7	11,4	13	13,0	13,0
P3	5,099	3		6,08	10,2	8,25	3,61	6,71	3	5	4,47	5,66	7,07	5	6,71	8,94	8,25	9,06	10,4	11	12,2	13,6	13,2
P4	2,236	3,16	6,08		4,12	2,24	9,06	2	6,32	4,24	8,54	10,4	4,12	7,21	10,3	7,28	8,06	10,6	12,7	11,7	13,6	12,2	12,6
P5	5,831	7,28	10,2	4,12		2	13	4,12	10,0	7,28	12,2	14,1	5,83	10,4	13,6	8,49	10	13,0	15,3	13,5	15,6	12,5	13,6
P6	4,243	5,39	8,25	2,24	2		11	2,24	8,06	5,39	10,2	12,2	4,24	8,54	11,7	7,21	8,49	11,4	13,6	12,0	14,1	11,7	12,5
P7	8,544	6,32	3,61	9,06	13	11		9,06	3,16	6,32	2,24	2,24	8,54	4,24	4	9,22	7,81	7,28	8	9,49	10,0	13,0	12,1
P8	4,123	4,24	6,71	2	4,12	2,24	9,06		6	3,16	8,06	10,0	2,24	6,32	9,49	5,39	6,40	9,22	11,4	10	12,0	10,2	10,8
P9	6,403	4,24	3	6,32	10,0	8,06	3,16	6		3,16	2,24	4,12	5,39	2	4,24	6,40	5,39	6,08	7,62	8	9,22	10,8	10,2
P10	5,385	4	5	4,24	7,28	5,39	6,32	3,16	3,16		5	7	2,24	3,16	6,32	4,12	4,12	6,40	8,49	7,62	9,43	9,06	9,06
P11	8,602	6,40	4,47	8,54	12,2	10,2	2,24	8,06	2,24	5		2	7,07	2,24	2,24	7,21	5,66	5,1	6,08	7,28	8	10,8	9,95
P12	10,3	8,06	5,66	10,4	14,1	12,2	2,24	10,0	4,12	7	2		9,06	4,12	2,24	8,94	7,21	5,83	6,08	8,06	8,25	12,0	10,8
P13	6	5,39	7,07	4,12	5,83	4,24	8,54	2,24	5,39	2,24	7,07	9,06		5	8,06	3,16	4,24	7,21	9,43	7,81	9,9	8,06	8,54
P14	7,810	5,83	5	7,21	10,4	8,54	4,24	6,32	2	3,16	2,24	4,12	5		3,16	5	3,61	4,12	5,83	6	7,28	8,94	8,25
P15	10,63	8,49	6,71	10,3	13,6	11,7	4	9,49	4,24	6,32	2,24	2,24	8,06	3,16		7,28	5,39	3,61	4	5,83	6,08	9,9	8,60
P16	9,055	8,06	8,94	7,28	8,49	7,21	9,22	5,39	6,40	4,12	7,21	8,94	3,16	5	7,28		2	5,1	7,28	5	7,21	5	5,39
P17	9,487	8,06	8,25	8,06	10	8,49	7,81	6,40	5,39	4,12	5,66	7,21	4,24	3,61	5,39	2		3,16	5,39	3,61	5,66	5,39	5
P18	11,66	9,85	9,06	10,6	13,0	11,4	7,28	9,22	6,08	6,40	5,1	5,83	7,21	4,12	3,61	5,1	3,16		2,24	2,24	3,16	6,40	5
P19	13,60	11,7	10,4	12,7	15,3	13,6	8	11,4	7,62	8,49	6,08	6,08	9,43	5,83	4	7,28	5,39	2,24		3,16	2,24	7,62	5,83
P20	13	11,4	11	11,7	13,5	12,0	9,49	10	8	7,62	7,28	8,06	7,81	6	5,83	5	3,61	2,24	3,16		2,24	4,47	2,83
P21	14,76	13	12,2	13,6	15,6	14,1	10,0	12,0	9,22	9,43	8	8,25	9,9	7,28	6,08	7,21	5,66	3,16	2,24	2,24		6,08	4,12
P22	14,04	13,0	13,6	12,2	12,5	11,7	13,0	10,2	10,8	9,06	10,8	12,0	8,06	8,94	9,9	5	5,39	6,40	7,62	4,47	6,08		2
P23	14,32	13,0	13,2	12,6	13,6	12,5	12,1	10,8	10,2	9,06	9,85	10,8	8,54	8,25	8,60	5,39	5	5	5,83	2,83	4,12	2	

Record: 14 of 23

Figura 12. Distanțele între punctele norului $\mathcal{P}_1 \subset \mathcal{R}^2$

CBD: REGIUNILE DE CĂUTARE

*	N	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
R																								
1	3	1	1		1																			
2	2	1	1																					
3	1			1																				
4	4	1			1		1		1															
5	2					1	1																	
6	4				1	1	1		1															
7	3							1				1	1											
8	4				1		1		1					1										
9	3									1		1			1									
10	2										1			1										
11	6							1		1		1	1		1	1								
12	4							1				1	1			1								
13	3								1		1			1										
14	3									1		1			1									
15	3											1	1			1								
16	2																1	1						
17	2																1	1						
18	3																	1	1	1				
19	3																	1	1		1			
20	3																	1		1	1			
21	3																		1	1	1			
22	2																						1	1
23	2																						1	1

Record: 24 of 24

Figura 13. Distribuția punctelor norului $\mathcal{P}_I \subset \mathcal{R}^2$ în regiunile de căutare

CBD: ZGOMOTELE ȘI GRUPURILE

*	N	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
Z	5			1												1	1						1	1
1	8	1	2		4	5	6		8		10			13										
2	6							7		9		11	12		14	15								
3	4																		18	19	20	21		

Figura 14. Gruparea punctelor din norul $\mathcal{P}_I \subset \mathcal{R}^2$: 3 clustere (1, 2, 3) și zgomotele (Z)

Figura 15 ilustrează norul de puncte \mathcal{P}_I înainte și după procesul de grupare a punctelor.

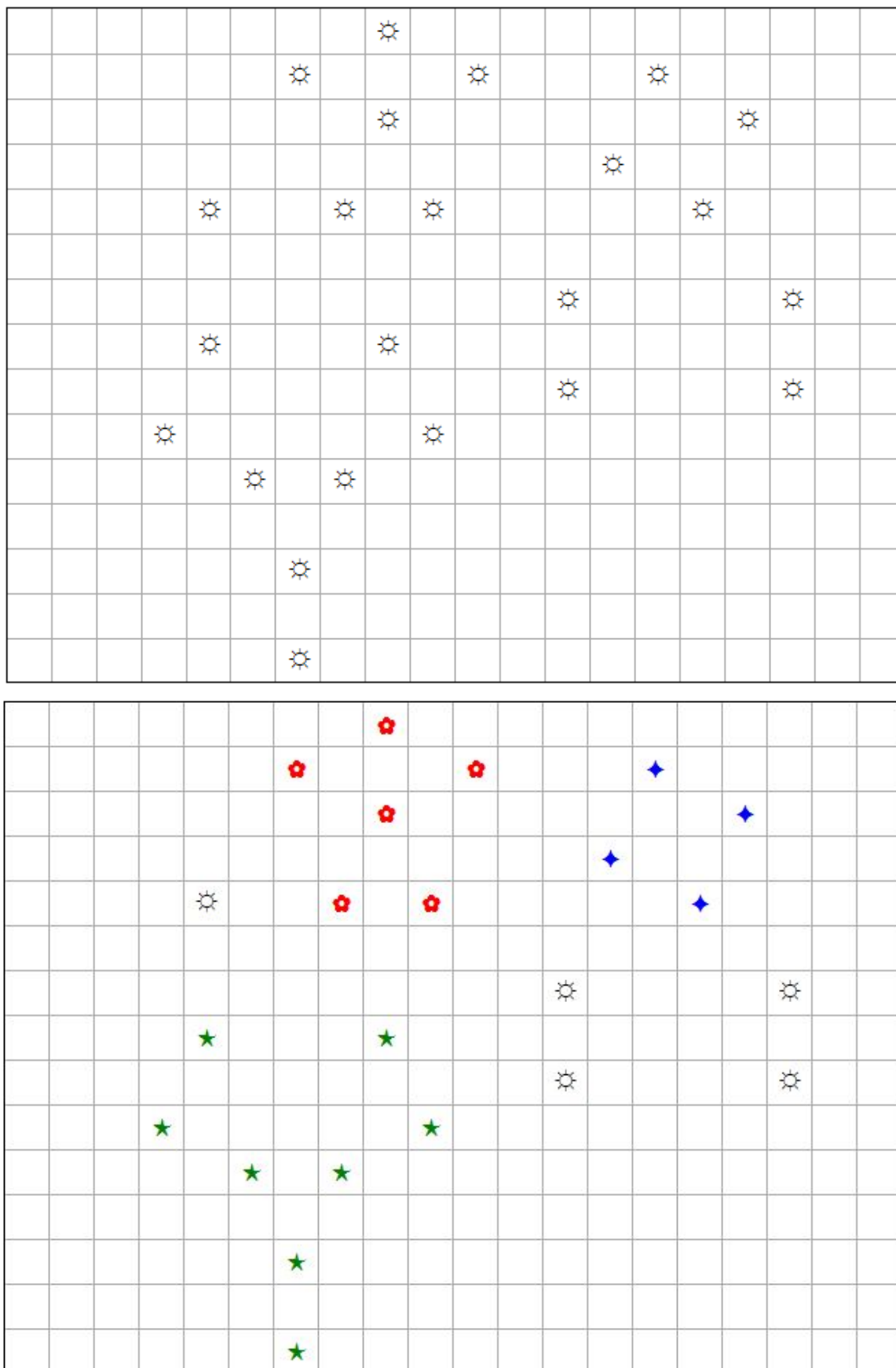


Figura 15. Norul $\mathcal{P}_I \subset \mathcal{R}^2$, clusterelor (🌺, ◆, ★) și zgomotele (☼)

Exemplul 2. Mulțimea $\mathcal{P}_2 \subset \mathcal{R}^8$ este formată din 15 reviste având asociate valori pentru un set de 8 indicatori bibliometrici [8]:

Journal	Total No of Cites	2-year IF	5-year IF	Immediacy Index	No. Of Articles	Cited Half-Life	Eigenfactor Score	Article Influence
1	44847	5,996	6,634	1,743	581	10,0	0,09218	1,979
2	24743	5,289	4,115	0,955	404	9,3	0,05119	1,213
3	6011	3,651	3,099	0,539	357	9,2	0,03615	0,967
4	5932	3,095	3,863	0,496	258	8,9	0,01553	0,864
5	4682	2,940	2,910	0,487	175	8,8	0,01422	0,824
6	4074	2,366	2,174	0,407	172	8,3	0,01404	0,715
7	3899	2,339	2,456	0,310	167	8,2	0,01289	0,709
8	3027	2,266	2,397	0,308	145	7,4	0,01069	0,629
9	2991	2,021	2,006	0,290	133	6,8	0,00872	0,584
10	2101	1,622	1,918	0,206	109	6,5	0,00622	0,547
11	1177	1,091	1,298	0,163	105	5,9	0,00483	0,381
12	956	1,049	1,497	0,156	86	5,0	0,00305	0,374
13	943	0,782	1,043	0,093	64	4,9	0,00251	0,298
14	443	0,778	1,017	0,057	62	4,8	0,00161	0,279
15	269	0,280	0,589	0,048	31	4,4	0,00058	0,180

Figura 16. Valori ale indicatorilor bibliometrici pentru un set de reviste

Valorile indicatorilor bibliometrici sunt normalizate și proiectate în intervalul $[0, 50]$, ele reprezintă coordonatele punctelor norului $\mathcal{P}_2 \subset \mathcal{R}^8$; $eps = 17$; $MinPts = 4$.

Coordonate	c1	c2	c3	c4	c5	c6	c7	c8
P1	50	50	50	27,4	34,8	46,5	50	50
P2	5,2	44,1	31	50	9,4	29,5	7,7	30,7
P3	1,1	30,4	23,4	14	30,7	24,5	19,6	21,8
P4	6,6	25,8	29,1	8,8	11,4	34	7,6	24,4
P5	27,6	24,5	21,9	14,2	50	44	27,8	20,8
P6	6,7	19,7	16,4	8,9	12,5	46	7	15,9
P7	4,3	19,5	18,5	15,5	22,2	22	8,4	17,9
P8	2,3	18,9	18,1	1,4	5,3	41,5	2,6	18,1
P9	3,4	16,9	15,1	11,7	14,4	25	5,8	14,8
P10	4,5	13,5	14,5	5,9	15,1	41	4,7	13,8
P11	3,3	9,1	9,8	4,7	14,8	50	3,4	9,5
P12	0,5	8,7	11,3	4,5	5,5	24	0,9	9,6
P13	1,3	6,5	7,9	1,6	9	37	1,7	7,5
P14	1,1	6,5	7,7	2,7	7,4	32,5	1,4	7
P15	0,3	2,3	4,4	8,3	2,7	44,5	0,3	4,5

Figura 17. Coordonatele punctelor norului $\mathcal{P}_2 \subset \mathcal{R}^8$

CBD: DISTANȚELE

*	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1	0,00	77,52	76,70	80,06	60,82	88,17	87,59	95,37	93,82	94,37	101,04	107,64	107,69	108,81	112,98
P2	77,52	0,00	47,52	45,85	67,90	54,78	48,34	59,31	52,58	60,11	68,17	65,35	70,26	69,46	72,31
P3	76,70	47,52	0,00	26,92	39,50	34,76	19,35	39,46	27,70	35,26	45,24	42,92	46,01	45,85	55,13
P4	80,06	45,85	26,92	0,00	50,35	20,40	22,49	20,00	21,89	23,68	34,44	32,61	35,17	35,46	42,87
P5	60,82	67,90	39,50	50,35	0,00	48,84	47,15	59,13	53,33	50,87	55,34	66,48	62,92	64,77	69,56
P6	88,17	54,78	34,76	20,40	48,84	0,00	27,02	13,24	21,81	9,87	16,13	28,54	22,82	25,27	27,59
P7	87,59	48,34	19,35	22,49	47,15	27,02	0,00	30,05	10,94	24,21	35,15	26,68	32,18	30,93	41,16
P8	95,37	59,31	39,46	20,00	59,13	13,24	30,05	0,00	22,28	13,65	20,35	23,34	20,15	21,80	26,85
P9	93,82	52,58	27,70	21,89	53,33	21,81	10,94	22,28	0,00	17,48	28,23	16,53	22,58	20,85	31,66
P10	94,37	60,11	35,26	23,68	50,87	9,87	24,21	13,65	17,48	0,00	12,07	21,55	14,94	17,48	22,86
P11	101,04	68,17	45,24	34,44	55,34	16,13	35,15	20,35	28,23	12,07	0,00	27,91	15,28	19,78	17,57
P12	107,64	65,35	42,92	32,61	66,48	28,54	26,68	23,34	16,53	21,55	27,91	0,00	14,55	10,21	23,61
P13	107,69	70,26	46,01	35,17	62,92	22,82	32,18	20,15	22,58	14,94	15,28	14,55	0,00	4,94	13,52
P14	108,81	69,46	45,85	35,46	64,77	25,27	30,93	21,80	20,85	17,48	19,78	10,21	4,94	0,00	15,30
P15	112,98	72,31	55,13	42,87	69,56	27,59	41,16	26,85	31,66	22,86	17,57	23,61	13,52	15,30	0,00

Record: 15 of 15

Figura 18. Distanțele între punctele norului $\mathcal{P}_2 \subset \mathcal{R}^8$

CBD: REGIUNILE DE CĂUTARE

*	N	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
R																
1	1	1														
2	1		1													
3	1			1												
4	1				1											
5	1					1										
6	4						1		1		1	1				
7	2							1		1						
8	3						1		1		1					
9	3							1		1			1			
10	5						1		1		1	1		1		
11	4						1				1	1		1		
12	4									1			1	1	1	
13	6										1	1	1	1	1	1
14	4												1	1	1	1
15	3													1	1	1

Record: 16 of 16

Figura 19. Distribuția punctelor norului $\mathcal{P}_2 \subset \mathcal{R}^8$ în regiunile de căutare

*	N	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
Z	7	1	1	1	1	1		1		1						
1	8						6		8		10	11	12	13	14	15
2																

Figura 20. Gruparea punctelor din norul $\mathcal{P}_2 \subset \mathcal{R}^8$: 1 cluster (1) și zgomotele (Z)

4. CONCLUZII

① Provocările cu care se confruntă un sistem suport pentru decizii de bibliotecă sunt: îmbunătățirea satisfacției utilizatorilor în sediul propriu și în afara lui, elaborarea unei politici de achiziție orientată către cerere; îmbunătățirea conservării colecțiilor; optimizarea fluxurilor și alocării resurselor umane și financiare; diversificarea și creșterea veniturilor; comunicarea mai bună cu partenerii.

② Domeniile de activitate ale bibliotecii, posibile de îmbunătățit, sunt: dezvoltarea colecțiilor; accesibilitatea colecțiilor; accesul la publicații; utilizarea bibliotecii; digitalizarea colecțiilor; serviciile bibliografice, asistența și îndrumarea; potențialul de dezvoltare; conservarea colecțiilor; managementul.

③ Utilizatorii sistemului solicită un spectru larg de expertize, de la căutări simple la statistici avansate. Pentru a putea adapta serviciile oferite de sistem la cerințele fiecărei categorii de utilizatori principalele categorii de servicii care trebuie avute în vedere sunt: căutările simple, care furnizează rapoarte predefinite și valori ale indicatorilor operaționali și/sau de performanță; interogările avansate și/sau personalizate; analizele avansate, care implică navigare multidimensională și funcții puternice de analiză; simulările și statisticile avansate.

④ Arhitectura necesară sistemului suport pentru decizii al unei biblioteci este o combinație dintre o tehnologie de management a rezolvatoarelor flexibile și o tehnologie de management a bazelor de date menită să asigure integrarea depozitării datelor cu rezolvatoarele analitice (prelucrare analitică on-line) și cu rezolvatoarele data mining (mineritul datelor și descoperirea cunoștințelor).

⑤ Pentru realizarea depozitului de date este foarte important ca proiectanții să urmeze o metodologie de proiectare conceptuală consolidată și robustă dat fiind că dezvoltarea acestuia este un proces foarte scump chiar în condițiile actuale când există instrumente software care oferă soluții prefabricate acoperind toate etapele din ciclul de viață al unui depozit de date.

Concepția și implementarea sistemului suport pentru decizii al unei biblioteci, ca de altfel ale oricărui sistem informatic, sunt influențate de către o serie de factori, între care pot fi menționați: obiectivele urmărite; personalul implicat; recomandările, normele și standardele utilizate; restricțiile impuse de către instituție; evoluția mediului; bugetul disponibil pentru realizare; termenele de finalizare.

⑥ Obiectivele sistemului suport pentru decizii de bibliotecă sunt:

↳ furnizarea de indicatori care să permită evaluarea în timp a conformității cu obiectivele bibliotecii prin: evaluarea rezultatelor obținute, sesizarea tendințelor,

alerte; evaluarea indicatorilor operaționali, de performanță și bibliometrici; rapoarte de activitate.

↳ furnizarea unor instrumente de analiză a tendințelor, de sesizare a situațiilor decizionale și de sugerare a unor acțiuni corespunzătoare (analize complexe, simulări, prognoze) în vederea luărilor de decizii.

↳ integrarea datelor și compararea informațiilor din aplicații informatice existente.

↳ simplificarea accesului la informație prin schimb transparent de informații și diseminare accelerată a informațiilor.

⑦ Pentru stabilirea cerințelor informaționale se impune aplicarea cu discernământ a prevederilor normative specifice domeniului bibliotecilor elaborate, recomandate și utilizate atât pe plan intern cât și pe plan internațional privind: descrierile bibliografice; indicatorii operaționali; indicatorii de performanță și indicatorii bibliometrici. Practic trebuie definit un sistem personalizat, unic, unitar, coerent și evolutiv de indicatori.

⑧ Pentru personalul implicat se impune următoarea structură: un comitet de management; o echipă de proiect; grupuri de lucru cu utilizatorii; o firmă de consultanță pentru analiza cerințelor; un subcontractant pentru dezvoltare-implementare.

⑨ La evaluarea resurselor financiare necesare trebuie avute în vedere: instrumentele de fundamentare a deciziilor (OLAP, Data mining ș.a.); metodele și instrumentele de realizare a depozitului de date, un instrument ETL pentru date ; un instrument de raportare; consultanța; contractul pentru dezvoltarea sistemului.

Pentru o bibliotecă avantajele majore ale sistemului suport pentru decizii sunt: asigură informații de calitate și noi instrumente de management; rezolvă faze tehnice critice privind furnizarea, modelarea și stocarea datelor; satisface cerințe tehnice actuale și viitoare; satisface cerințele utilizatorilor; este adaptabil; susține trecerea la o cultură orientată către performanță și impune personalului dezvoltarea în consecință a abilităților; promite mari îmbunătățiri în modul de înțelegere a ceea ce se face în prezent și a ceea ce se preconizează pentru viitor.

BIBLIOGRAFIE

- [1] ATZENI, P.; CAPPELLARI, P.; TORLONE, R.; BERNSTEIN, P.A.; GIANFORME, G. (2008) *Model independent schema translation*. In Journal of Very Large Data Bases, 17, pp 1347–1370.
- [2] BONIFATI, A.; CATTANEO, F.; CERI, S.; FUGGETTA, A.; PARABOSCHI, S. (2001) *Designing data marts for data warehouses*. In ACM Transactions on Software Engineering and Methodology, 10, pp 452–483.
- [3] BORNE, P.; POPESCU, P.; FILIP, F.-G.; STEFANOIU, D. (2013) *Optimisation en sciences de l'ingénieur – méthodes exactes*, Paris: Lavoisier, 332 p.
- [4] BNF - Bibliothèque nationale de France (2012) *Fonctionnalités requises des notices bibliographiques : rapport final*. Trad. de : „Functional requirements for bibliographic records : Final Report” - 2e édition française, Paris, BNF, 100 p.
- [5] BURSTEIN, F.; HOLSAPPLE, C.W. (Eds) (2008), *Handbook on Decision Support Systems 1 : Basic Themes*, International Handbooks on Information Systems, Springer-Verlag, 854 p.
- [6] BURSTEIN, F.; HOLSAPPLE, C.W. (Eds), (2008), *Handbook on Decision Support Systems 2 : Variations*, International Handbooks on Information Systems, Springer-Verlag, 800 p.
- [7] CABIBBO, L.; TORLONE, R. (1998) *A logical approach to multidimensional databases*. In 6th International Conference on Extending Database Technology (EDBT'98), Springer-Verlag, pp 183-197.
- [8] DURIEUX, V.; GEVENOIS, P. A. (2010) Bibliometric indicators: quality measurements of scientific publication. In RADIOLOGY, 255 (2), pp 342-351.
- [9] FILIP, F.-G. (2005) *Decizie asistată de calculator: decizii, decidenți - metode de bază și instrumente informatice asociate*, Ed. a 2-a, rev., București: Ed Tehnică, 376 p.
- [10] FILIP, F.-G. (2007) *Sisteme suport pentru decizii*, Ed. a 2-a, rev., București: Editura Tehnică, 364 p.
- [11] GIORGINI, P.; RIZZI, S.; GARRETTI, M. (2008) *GRAnD: a goal-oriented approach to requirement analysis in data warehouses*. In Decision Support Systems, 45, pp 4–21.
- [12] GOLFARELLI, M.; MAIO, D.; RIZZI, S. (1998) *Conceptual design of data warehouses from E/R schema*. In Thirty-First Annual Hawaii International Conference on System Sciences, 7, pp. 334–344.
- [13] GOLFARELLI, M.; RIZZI, S. (2009) *Data Warehouse Design: Modern Principles and Methodologies*, McGraw-Hill, 445 p.
- [14] HIRSCH, J. E. (2005) *An index to quantify an individual's scientific research output* (<http://arxiv.org/pdf/physics/0508025.pdf>)
- [15] IFLA - International Federation of Library Associations and Institutions (2008) *Functional requirements for bibliographic records : Final Report*. (http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf)
- [16] INS - Institutul National de Statistica (2012) *Cult1 – Activitatea bibliotecilor. În Chestionare statistice, Statistica culturii*. (<http://www.insse.ro/cms/files/chestionare/cult/CULT1%202012.pdf>)
- [17] ISO (2009) *TR28118 Information and documentation – Performance indicators for national libraries*. (http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=01&ics2=140&ics3=20&csnumber=44512)
- [18] ISO (2013) *2789 Information and documentation - International library statistics* (<http://webstore.ansi.org/RecordDetail.aspx?sku=ISO%202789:2013&source=google&adgroup=iso13&gclid=CLb-2s78kLsCFQ1c3godxDcAeQ>)
- [19] JOUGUELET, S.; PARDÉ, T. (2005) *Defining and building a decision support system: the experience of the Bibliothèque nationale de France*. In LIBER QUARTERLY The Journal of the Association of European Research Libraries, 15, No 3/4.
- [20] LECHTENBÖRGER, J.; VOSSSEN, G. (2003) *Multidimensional normal forms for data warehouse design*. In Information Systems, 28, pp 415–434.
- [21] LOR, P. J. (1997) Guidelines for legislation for national library services CH–97/WS/7.

- [22] LUJAN-MORA, S.; TRUJILLO, J.; SONG, I.Y. (2006) *A UML profile for multidimensional modelling in data warehouses*. In *Data & Knowledge Engineering*, 59, pp 725–769.
- [23] MAZÓN, J.-N.; TRUJILLO, J.; SERRANO, M.; PIATTINI, M. (2005) *Designing data warehouses: from business requirement analysis to multidimensional modeling*. In: K. Cox, E. Dubois, Y. Pigneur, S.J. Bleistein, J. Verner, A.M. Davis, R. Wieringa (Eds.), *Requirements Engineering for Business Need and IT Alignment*, Paris, pp. 44–53.
- [24] MAZÓN, J.-N.; TRUJILLO, J.; LECHTENBÖRGER, J. (2006) *A Set of QVT Relations to assure the correctness of data warehouses by using multidimensional normal forms*. In: D.W. Embley, A. Olivé, S. Ram (Eds.), *ER: Lecture Notes in Computer Science*, vol. 4215, *Conceptual Modeling*, pp. 385–398.
- [25] MAZÓN, J.-N.; TRUJILLO, J.; LECHTENBÖRGER, J. (2007) *Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms*. In *Data & Knowledge Engineering*, 63, pp 725–751.
- [26] MAZÓN, J.-N.; PARDILLO, J.; SOLER, E.; GLORIO, O.; TRUJILLO, J. (2008) *Applying the i* framework to the development of data warehouses*. In *Proceedings of the 3rd International i* Workshop (iStar'08, February 11-12)*, pp 79 – 82.
- [27] MAZÓN, J.-N.; TRUJILLO, J. (2009) *A hybrid model driven development framework for the multidimensional modeling of data warehouses*, In *SIGMOD Record*, 38, pp 12–17.
- [28] PENDLEBURY, D.A. (2008) *Using Bibliometrics in Evaluating Research* (http://wokinfo.com/media/mtrp/UsingBibliometricsinEval_WP.pdf)
- [29] PHIPPS, C.; DAVIS K.C. (2002) *Automating data warehouse conceptual schema design and evaluation*. In: Laks V.S. Lakshmanan (Ed.), *DMDW: CEUR Workshop Proceedings*, 58, *Design and Management of Data Warehouses*, pp 23–32.
- [30] RAFANELLI, M.(2003)*Multidimensional databases: problems and solutions*. Idea Group Inc., 446 p.
- [31] REHN, C.; KRONMAN, U.; WADSKOG, D. (2007) *Bibliometric indicators – definitions and usage at Karolinska Institutet*. (http://ki.se/content/1/c6/01/79/31/Bibliometric%20indicators%20-%20definitions_1.0.pdf)
- [32] ROMERO, O.; ABELLÓ, A. (2009) *A survey of multidimensional modeling methodologies*, In *International Journal of Data Warehousing and Mining*, 5, pp 1–23.
- [33] ROMERO, O.; ABELLÓ, A. (2010) *Automatic validation of requirements to support multidimensional design*. In *Data & Knowledge Engineering*, 69, pp 917–942.
- [34] SCHNEIDER, M. (2008) *A general model for the design of data warehouses*. In *International Journal of Production Economics*, 112, pp 309–325.
- [35] STEFANOV, V.; LIST, B. (2007) *A UML profile for modeling data warehouse usage*. In: J. Hainaut et al. (Eds.), *ER Workshops: Lecture Notes in Computer Science*, vol. 4802, *Advances in Conceptual Modeling – Foundations and Applications*, pp. 137–147.
- [36] TÎRZIMAN, E. (2000) *Procesele de bibliotecă: abordare în contextul utilizării noilor tehnologii*. București : Editura Universității București, 108 p.
- [37] TRIA (DI), F.; LEFONS, E.; TANGORRA, F. (2012) *Hybrid methodology for data warehouse conceptual design by UML schemas*. In *Information and Software Technology*, 54, pp 360–379.
- [38] WEST, J.; ALTHOUSE, B.; ROSVALL, M.; BERGSTROM, C.; BERGSTROM, T. (2008) *Eigenfactor Score and Article Influence Score: Detailed methods*. *Methods* version 2.01. (<http://www.eigenfactor.org/methods.pdf>).
- [39] * * * (2011) *The evolution of journal assessment - SNIP & SJR new perspectives in journal metrics*. (http://www.journalmetrics.com/documents/Journal_Metrics_Whitepaper.pdf).
- [40] * * * (2013) <http://en.wikipedia.org/wiki/DBSCAN> .

LISTA FIGURILOR

- Figura 1. Algoritm de generare a arborilor de atribute
Figura 2. Arhitectura sistemului suport pentru decizii
Figura 3. Dimensiuni, niveluri dimensionale și căi de agregare
Figura 4. Cub de date privind serviciile bibliotecare
Figura 5. Cub de date privind aparițiile editoriale
Figura 6. Cub de date privind calitatea publicărilor
Figura 7. Schema „constelație” a depozitului de date
Figura 8. Algoritm de generare nucleu de autoritate
Figura 9. Algoritm de clusterizare bazată pe densitate
Figura 10. Meniul programului CBD
Figura 11. Coordonatele punctelor norului $\mathcal{P}_1 \subset \mathbb{R}^2$
Figura 12. Distanțele între punctele norului $\mathcal{P}_1 \subset \mathbb{R}^2$
Figura 13. Distribuția punctelor norului $\mathcal{P}_1 \subset \mathbb{R}^2$ în regiunile de căutare
Figura 14. Gruparea punctelor din norul $\mathcal{P}_1 \subset \mathbb{R}^2$: 3 cluster (1, 2, 3) și zgomotele (Z)
Figura 15. Norul $\mathcal{P}_1 \subset \mathbb{R}^2$, cluster (★, ◆, ◆) și zgomotele (⊗)
Figura 16. Valori ale indicatorilor bibliometrici pentru un set de reviste
Figura 17. Coordonatele punctelor norului $\mathcal{P}_2 \subset \mathbb{R}^8$
Figura 18. Distanțele între punctele norului $\mathcal{P}_2 \subset \mathbb{R}^8$
Figura 19. Distribuția punctelor norului $\mathcal{P}_2 \subset \mathbb{R}^8$ în regiunile de căutare
Figura 20. Gruparea punctelor din norul $\mathcal{P}_2 \subset \mathbb{R}^8$: 1 cluster (1) și zgomotele (Z)

LISTA TABELELOR

- Tabelul 1. Lista indicatorilor operaționali ai bibliotecii
Tabelul 2. Obiectivele bibliotecii și indicatorii de performanță
Tabelul 3. Produsele unei activități intelectuale sau artistice
Tabelul 4. Responsabilii pentru produsele unei activități intelectuale sau artistice
Tabelul 5. Subiectele produselor unei activități intelectuale sau artistice
Tabelul 6. Relațiile dintre entitățile bibliografice
Tabelul 7. Indicatori bibliometrici de productivitate
Tabelul 8. Indicatori bibliometrici de performanță (1 – 3)
Tabelul 9. Descrierea bibliografică a unei reviste
Tabelul 10. Descrierea bibliografică a unui *proceedings*
Tabelul 11. Descrierea bibliografică a unei monografii
Tabelul 12. Descrierea bibliografică a unei monografii, format UNIMARC
Tabelul 13. Descrierea bibliografică a unui articol de revistă
Tabelul 14. Descrierea bibliografică a unui capitol dintr-o monografie
Tabelul 15. Maparea pe sursele de date a descrierii bibliografice - document
Tabelul 16. Maparea pe sursele de date a descrierii bibliografice - publicație