

CRIS - SISTEME DE INFORMARE PRIVIND ACTIVITATEA DE CERCETARE

Gabriel Neagu

Ileana Trandafir

Ana-Maria Borozan

Mădălina Zamfir

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

Rezumat. Utilizarea eficientă a fondurilor de cercetare, valorificarea potențialului uman și de infrastructură existent, susținerea orientării pe proiecte integrate, bazate pe colaborare la nivel național și internațional, promovarea rezultatelor proiectelor de cercetare evidențiază necesitatea unor soluții informatice avansate, de administrare a informației despre activitatea de cercetare. Lucrarea de față prezintă succint două concepte esențiale, promovate la nivelul Comisiei Europene, ca rezultat al eforturilor demarate în urmă cu două decenii, în domeniul bazelor de date pentru activitatea de cercetare: CRIS – Current Research Information Systems și CERIF – Common European Research Information Format.

Cuvinte cheie: activitatea de cercetare, sistem de informare, CRIS, ERGO, euroCRIS, ciclul de viață, structuri de date, CERIF, modelul complet de date, XML, formatul de schimb de date.

1. Introducere

Administrarea performantă a informației referitoare la activitatea de cercetare are o influență determinantă asupra eficienței acestei activități. Facilitarea localizării expertizei în diverse domenii și identificarea posibilităților de colaborare, limitarea paralelismelor în abordarea unor subiecte de cercetare și fundamentarea corectă a cererilor de finanțare, furnizarea de informație statistică de mare utilitate pentru managementul acestei activități pe toate nivelurile decizionale reprezintă argumente importante în acest sens.

Începuturile dezvoltării sistemelor de administrare a informației privind activitatea de cercetare, cunoscute sub numele de *CRIS - Current Research Information Systems*, datează din anii '70. În prima etapă, sistemele CRIS au acoperit numai informația despre proiecte de cercetare, urmărind să răspundă cu prioritate interesului imediat al cercetătorilor. Ulterior, a devenit clar că interesul pentru serviciile furnizate de aceste sisteme este mult mai larg, ceea ce a dus la diversificarea informației administrate de aceste sisteme: programe de cercetare, laboratoare și colective de cercetare, centre și institute de profil, personal de cercetare, expertiză și competențe, surse de finanțare, rezultate ale cercetărilor (publicații, patente, produse), infrastructura de cercetare (echipamente, instalații), evenimente științifice, baze de date privind transferul de tehnologie, standarde în domenii tehnologice de vârf, proceduri și reglementări privind proprietatea intelectuală. Acest proces continuă și în prezent, prioritare fiind informațiile privind utilizarea rezultatelor și evaluarea activității de cercetare.

Un salt important în evoluția CRIS s-a produs o dată cu dezvoltarea serviciilor Internet, crescând astfel potențialul de diversificare a modalităților de utilizare a datelor, precum și o mai bună motivare a furnizorilor de informații.

Indiferent de etapa dezvoltării acestor sisteme, o problemă majoră, care a preocupat comunitatea științifică implicată în această activitate, a constituit-o convergența către o soluție unitară pentru structurarea datelor, care să asigure răspuns la două cerințe esențiale: compatibilitatea bazelor de date respective și facilitarea schimbului de informații între sisteme.

Unul din rezultatele notabile al acestor preocupări pe plan european l-a constituit dezvoltarea formatului unitar de date *CERIF – Common European Research Information Format*, care a cunoscut, până în prezent, trei versiuni de referință. CERIF are ca obiectiv stimularea accesului la informația despre potențialul și competențele de cercetare din spațiul european, pentru valorificarea lor superioară.

Prezenta lucrare este dedicată problematicii CRIS și prezentării în acest context a formatului CERIF.

2. Evoluția sistemelor CRIS

2.1. Scurt istoric

Atenția acordată pe plan internațional sistemelor de informare privind activitatea de cercetare (CRIS) nu este de dată recentă. În 1970, a fost publicat, sub egida UNESCO, un manual referitor la potențialul științific și tehnologic național și reflectarea acestuia în sistemele pentru cercetare-dezvoltare [1]. Acest document descria metodologia de evaluare a potențialului științific la nivel național și analiza posibilitatea de utilizare a informației științifice pentru formularea unei politici naționale în domeniul științei.

În 1971, UNISIST a publicat un studiu de fezabilitate privind realizarea Sistemului Mondial de Informare

Științifică [2]. Studiul, elaborat tot sub egida UNESCO, recomanda UNISIST să stimuleze cooperarea internațională între sistemele de informare și să îmbunătățească instrumentele utilizate în acest sens.

În 1972, articolul [3] a prezentat sistemul de schimb al informației științifice *SSIE- Smithsonian Science Information Exchange*, care conținea informație extinsă referitoare la proiectele științifice, și se baza pe o infrastructură evoluată pentru acea perioadă (terminale video, memorie de mare capacitate).

În 1975, UNISIST a publicat Ghidul inventarierii pe plan național a proiectelor de cercetare-dezvoltare în desfășurare [4]. În același an, pe baza documentelor [1] și [4], UNESCO a publicat un inventar pe plan mondial al serviciilor de informare privind activitatea curentă de cercetare [5].

Pe plan european, un eveniment important în dezvoltarea acestui domeniu l-a constituit workshop-ul pentru baze de date în cercetare, organizat la Bruxelles în perioada 30 septembrie - 2 octombrie 1987. Scopul acestui workshop a fost să ofere Comisiei Europene oportunitatea de a identifica problemele care trebuie rezolvate în vederea realizării unei rețele de informare științifică și tehnologică. Una din consecințele acestui workshop a fost înființarea *Grupului de lucru european pentru baze de date în cercetare*, care a recomandat utilizarea CERIF. Opțiunea a fost justificată de obiectivele majore ale acestui format:

- a. să permită schimbul de înregistrări cu informație despre proiectele de cercetare între sistemele din diferite țări membre ale comunității europene;
- b. să servească drept referință pentru proiectarea bazelor de date din cercetare.

Prima versiune de referință a CERIF a fost adoptată în anul 1991. Ulterior, soluția CERIF a fost oficial recomandată spre utilizare țărilor membre de către Comisia Europeană, ca format standard de informare privind proiectele de cercetare [6]. Documentul recomanda adoptarea progresivă a următoarelor măsuri, la nivelul țărilor membre:

- armonizarea bazelor de date la nivel național în domeniul cercetării și dezvoltării tehnologice;
- inventarierea proiectelor de cercetare-dezvoltare derulate în instituțiile de cercetare și în industrie în conformitate cu standardul CERIF.

Următoarea reuniune importantă la nivel european în domeniul bazelor de date pentru cercetare, a avut loc în 1992 și a vizat, în principal, schimbul de experiență în implementarea CERIF. Raportul de concluzii al acestei întruniri a fost transmis Comitetului pentru Cercetarea Științifică și Tehnică, în calitate de consultant al Comisiei Europene. Principala recomandare a acestui raport era lansarea unui proiect care să permită accesul dintr-un singur punct (one stop shop) la bazele de date compatibile CERIF din întreaga Comunitate Europeană.

În baza acestei recomandări, a fost constituit *Grupul de lucru ERGO (European Research Gateways On-line)*, care a primit mandatul de a investiga fezabilitatea acestui proiect. Dintre alternativele de soluții propuse de acest grup de lucru, Comitetul pentru Programul de Inovare al Comisiei Europene a decis în 1996 lansarea proiectului pilot ERGO.

2.2. Proiectul pilot ERGO

Proiectul a demarat efectiv în anul 1998, având ca obiectiv facilitarea accesului la bazele de date naționale pentru cercetare-dezvoltare, în vederea difuzării cât mai largi a informației privind proiectele de cercetare-dezvoltare din Europa. Prezentarea acestui proiect a constituit tema principală a conferinței *CRIS '98 - The way to innovation* [7], [8], [9].

Proiectul colecta într-un nod central informația de bază în limba engleză despre proiectele de cercetare din întreaga Europă. Această informație putea fi completată cu detalii preluate din bazele de date naționale sau prin persoana de contact pentru proiectul respectiv. Informația era oferită gratuit de către serviciile de informare regionale sau naționale, care transmiteau regulat rezumate privind conținutul bazei de date. Întreg proiectul se afla sub coordonarea Comisiei Europene. Proiectul era monitorizat de *ERGO Implementation Group (EIG)*, ai cărui membri erau furnizori de conținut din țările participante și țările asociate. Activitatea EIG era asistată de echipa centrală ERGO având în responsabilitate colectarea și prelucrarea datelor, precum și administrarea și documentarea site-ului web, care găzduia catalogul sistemului.

Obiectivele proiectului au fost următoarele:

- facilitarea accesului la serviciile naționale de informare științifică și tehnică;
- identificarea principalelor surse de informare și evaluarea posibilităților de acces și de utilizare a acestor surse la nivel european;

- dezvoltarea unor conexiuni și a unor instrumente care să permită accesul în mod coordonat la aceste surse, din țările membre ale Comunității Europene.

Utilizatorii țintă ai rezultatelor acestui proiect au fost: cercetători din institutele și centrele de profil, centrele de cercetare și inovare din industrie, organizații de consultanță în dezvoltarea științifică și tehnologică, decidenți în domeniul politicilor de cercetare-dezvoltare, intermediari de servicii informaționale și, într-o oarecare măsură, mass-media.

Soluția tehnică propusă de grupul de lucru s-a bazat pe modelul de catalog, model caracterizat prin simplitate la nivel central și complexitate pe nivelul distribuit al surselor primare de informare (figura 1).

Avantajele acestei soluții erau următoarele:

- schimburile informaționale între bazele de date sursă și nivelul central erau transparente utilizatorului;
- modelul susținea conceptul de piață deschisă pentru servicii de informare, unde deținătorii bazelor de date pot decide ce oferă și la ce preț, iar utilizatorii pot alege serviciile pentru care doresc să plătească;
- deținătorul de baze de date nu era afectat de modificările în interfața cu utilizatorul deoarece singura lui relație cu nivelul central o reprezintă exportul de date, conform cererilor primite de pe acest nivel;
- puteau fi implementate politici locale de finanțare și de asigurare a securității surselor de date negociate cu deținătorii bazelor de date la nivel național.

Proiectul a fost construit în două etape. În prima etapă, au fost avute în vedere aproximativ 30 de baze de date de proiecte cu informații prezentate în diverse limbi. În cea de-a doua etapă, aria de cuprindere privind informația referitoare la activitatea de cercetare s-a extins, incluzând organizații, expertiză, programe de cercetare și rezultate ale cercetărilor. De asemenea, a fost asigurat acces și altor clase de utilizatori decât cei conectați prin interfață web.

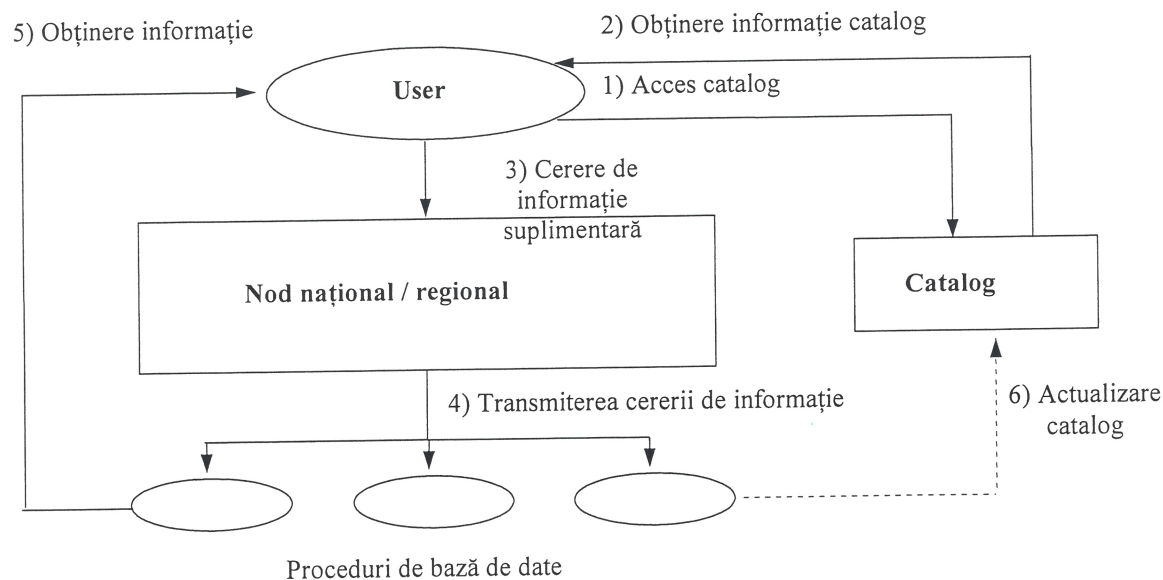


Figura 1. Modelul Catalog implementat în proiectul ERGO

Proiectul pilot ERGO a fost găzduit de site-ul CORDIS. ERGO a fost un proiect de succes, reușind să colecteze date despre peste 90.000 de proiecte naționale de C&D. Cu toate acestea, din motive procedurale, ERGO nu a putut fi continuat dincolo de faza de proiect pilot.

2.3. Asociația euroCRIS

EuroCRIS este o asociație non-profit recunoscută la nivel european și parțial în SUA ca o autoritate și un punct de referință în problematica CRIS: baze de date despre expertiza de cercetare, proiecte, instituții, infrastructură de cercetare și produse (inclusiv publicații), standarde, practici exemplare și ghiduri. Asociația a fost fondată în 1991 la Bergen, Norvegia, este marcă înregistrată și este prezentă pe Internet (www.eurocris.ro).

Principalul obiectiv al euroCRIS este să reprezinte un forum al organizațiilor și persoanelor interesate în schimbul de idei și în rezolvarea aspectelor legate de utilizarea tehnologiei informației în domeniul sistemelor de informare pentru cercetare. Dintre dezvoltările tehnologice mai recente în domeniu,

euroCRIS este interesat de utilizarea bazelor de date avansate, soluții de administrare a fluxurilor de activități, interfețe evaluate om-calculator, tehnologii Grid. Întreaga sa activitate este destinată îmbunătățirii disponibilității informațiilor despre cercetare.

Membrii euroCRIS sunt cercetători și practicieni în acest domeniu, cadre didactice și specialiști din administrația publică, centre informatice din mediul industrial și centre naționale de cercetare, precum și parteneri strategici pentru comunitatea de cercetare la nivel european: *ALLEA – All European Academies*, *CODATA - Committee on Data for Science and Technology*, *International Council for Science*, *EARMA - European Association of Research Managers & Administrators*, *ERCIM - European Research Consortium for Informatics and Mathematics*, *ESF - European Science Foundation*.

Asociația este custodele standardului CERIF, care asigură interoperabilitatea sistemelor CRIS și respectarea recomandărilor UE pentru țările membre.

Domeniile de interes major pentru activitatea euroCRIS sunt următoarele:

- bazele de date ale cercetării, orientate tematic sau specializate pe tipuri de informații (expertiză, proiecte, instalații și produse ale cercetării, inclusiv publicații);
- date specifice CRIS: seturi de date științifice, depozite de date instituționale (în regim Open Acces), sisteme de colectare a datelor, baze de cunoștințe și procese bazate pe sisteme workflow;
- întreținerea și diseminarea standardului CERIF în vederea asigurării interoperabilității sistemelor din familia CRIS;
- mecanisme de acces și interschimb de date, standarde și cele mai bune practici pentru dezvoltarea și implementarea CRIS.

EuroCRIS organizează întâlniri bianuale ale membrilor săi, workshop-uri, seminarii anuale, conferințe internaționale bienale; întreține forumuri de discuții on line și editează publicații în domeniul său de specialitate.

Conferințele euroCRIS pot să decidă cu privire la organizarea unor grupe de lucru, în cazul în care este nevoie de un efort focalizat pentru examinarea anumitor aspecte, pentru planificarea unor noi activități sau pentru investigarea pe termen scurt a unor probleme. Pot fi membre într-un grup de lucru orice organizație sau persoana membră euroCRIS cu expertiză legată de obiectul grupei de lucru. Câteva exemple:

- a. *CERIF – Best practices*: maximizarea interoperabilității și funcționalității sistemelor de informare pentru cercetare prin crearea și întreținerea standardului CERIF pentru sistemele CRIS;
- b. *Projects*: concentrarea expertizei tehnice euroCRIS în domeniul CRIS în vederea elaborării de propuneri și identificării de fonduri prin participarea la licitații, întreținerea și perfecționarea portalurilor de cercetare și a sistemelor integrate la nivel european în beneficiul dezvoltării tehnologice și a politicii cercetării;
- c. *External communications*: stimularea și promovarea comunicării în cadrul asociației și cu comunitatea globală CRIS, pregătirea de materiale promoționale, organizarea seminarului anual euroCRIS, întreținerea listelor de mail ale asociației.

3. Specificitatea sistemelor CRIS

3.1. Caracteristicile CRIS

Caracteristicile sistemelor CRIS sunt evidențiate pe baza unor cerințe specifice de utilizare, definiții privind profilul funcțional și rezultate din practica de realizare a acestei clase de sisteme:

a) Principalele categorii de utilizatori și interesele acestora:

- *instituții de cercetare*: analiza propriei activități, compararea cu instituțiile similare, fundamentarea unor politici și acțiuni manageriale pe termen scurt și mediu, dezvoltarea de colaborări și constituirea de consorții pe proiecte, stabilirea unor soluții de alimentare cu date din sistemul informatic propriu;
- *instituții guvernamentale*, cu responsabilități privind: orientarea, planificarea și coordonarea activității de cercetare la nivel departamental sau național, monitorizarea procesului de desfășurare a programelor naționale de cercetare, evaluarea de ansamblu a rezultatelor, elaborarea de analize statistice privind activitatea de cercetare;
- *utilizatori individuali*: cercetători interesați în identificarea unor realizări de referință în aceste domenii și evaluarea oportunității unor posibile colaborări, experți implicați în evaluarea și monitorizarea proiectelor de cercetare, reprezentanți ai instituțiilor guvernamentale, implicați în coordonarea și analiza desfășurării

programelor de cercetare, reprezentanți ai mediului de afaceri interesați în cofinanțarea unor proiecte și / sau preluarea rezultatelor activității de cercetare, reprezentanți mass-media.

b) Conținutul informațional al sistemului:

Informațiile considerate esențiale pentru satisfacerea cerințelor de informare ale categoriilor de utilizatori, definite mai sus, se referă, în mod esențial, la următoarele entități nucleu:

- *Proiecte* – titlul și informațiile asociate fiecărui proiect de cercetare individual;
- *Programe* - titlul și informațiile asociate unui program de cercetare în ansamblul său;
- *Expertiza* – cunoștințele și experiența cercetătorilor individuali;
- *Rezultate* – ieșirile proiectelor sau programelor de cercetare (produse, brevete, distincții);
- *Publicații* – o categorie specifică de rezultate ale activității de cercetare, constând din documentația rezultată din această activitate (rapoarte de etapă cu caracter public, sinteze, articole);
- *Organizații* – instituții sau organisme însărcinate cu finanțarea cercetării sau care derulează proiecte / programe de cercetare.

c) Implementarea de soluții eficiente privind achiziția datelor:

Aceste soluții trebuie să acopere ambele faze ale procesului de achiziție: furnizarea inițială a datelor și actualizările ulterioare. Ele implică nu numai aspecte procedurale, cât și tehnice. Astfel, din punct de vedere administrativ, sunt necesare:

- stabilirea din fazele inițiale ale proiectului a contactelor cu furnizorii / sursele de informații;
- negocierea cu fiecare furnizor a cerințelor de informații;
- încheierea unor acorduri de colaborare cu aceștia privind: conținutul informațional și formatul de preluare a datelor, termenele de livrare, procedurile de verificare și de control al datelor, procedurile de corecție și de tratare a excepțiilor, clauzele privind publicarea informației și respectarea drepturilor de autor.

Sub aspect tehnic, preluarea informației constă în primul rând din verificarea calității și a formatului acesteia, urmată de conversia și codificarea datelor, validarea conformității acesteia cu structura bazei de date CRIS (de ex. completitudinea în sensul respectării caracterului obligatoriu al câmpurilor, corectitudinea codificării datelor, asignarea corectă a cuvintelor cheie pe baza unui tezaur).

d) Alte cerințe funcționale:

- administrarea eficientă a conținutului informațional: completitudinea, coerența și disponibilitatea informațiilor;
- interactivitate și flexibilitate în formularea cererilor de regăsire informații;
- comanda on-line a unor documente;
- interoperabilitate cu alte baze de date, în relație de beneficiari ai conținutului informațional al CRIS;
- legături cu documentele full text și obiectele multimedia referite în baza de date;
- facilități de analiză a datelor, generarea de rapoarte statistice;
- tutorial on-line;
- administrarea informației multilingve.

În mod evident, facilitățile de căutare și navigare pe conținutul informațional al sistemului sunt determinante pentru utilizabilitatea sistemului. În acest sens, sunt considerate esențiale posibilitățile de căutare pe text (în cazul unor câmpuri semnificative), căutarea bazată pe clasificările atașate unor câmpuri indexate, furnizarea unor criterii predefinite de căutare, dar și posibilitatea utilizatorului de a-și defini propriile criterii de căutare, posibilitatea revenirii în căutările multinivel și navigări.

3.2. Ciclul de viață CRIS

Specificitatea evidențiată anterior se reflectă și în ciclul de viață CRIS, structurat conform etapelor prezentate în tabelul 1 și în figura 2 [10]. Față de ciclul de viață standard al unui sistem informatic, se remarcă: atenția acordată identificării surselor de informații și disponibilității acestora, orientarea analizei de sistem pe stabilirea conținutului informațional funcție de cerințele utilizatorilor, proiectarea centrată pe soluțiile de căutare și navigare, tratarea într-o etapă distinctă a proceselor de achiziție a datelor de intrare, accentul pus pe promovarea sistemului și stabilirea planului de diseminare la nivelul utilizatorilor potențiali.

Tabelul 1. Etapele ciclului de viață CRIS

Etapa	Activități, rezultate
Concept	Studiu de piață Analiza cost/beneficii, model economic Identificarea categoriilor de utilizatori potențiali Analiza disponibilității datelor și a restricțiilor de copyright
Propunere	Definirea scopului Identificarea utilizatorilor Definirea conținutului
Proiectare	Specificațiile bazei de date Structură și prezentare Clasificări și indexare Căutare și navigare
Achiziție și procesare informații	Planificarea colectării datelor Ghiduri privind colectarea datelor și controlul calității Preluarea datelor de intrare
Publicare/distribuție și marketing	Strategia de promovare a CRIS Plan de diseminare
Întreținere	

4. Formatul CERIF – o soluție unitară de structurare a datelor pentru CRIS

4.1. Prezentare generală

CERIF este întreținut și promovat de Asociația euroCRIS, în contextul responsabilităților acesteia privind dezvoltarea sistemelor CRIS.

Principalele obiective ale proiectării CERIF au fost următoarele:

- furnizarea unui model de date CRIS complet, care să asigure sistemelor CRIS existente flexibilitatea de a se mapa pe această structură de date;
- furnizarea unor soluții de bază pentru schimbul de date.

În stabilirea acestor obiective, s-a avut în vedere diversitatea structurilor adoptate de CRIS existente, ceea ce făcea practic imposibilă construirea unui model pentru schimbul de date care să satisfacă toate cerințele de acest gen.

Așa cum s-a menționat anterior, prima versiune a modelului de date CERIF a fost adoptată în anul 1991 și se referea numai la structura datelor privind proiectele de cercetare.

Următoarea versiune - CERIF 2000 - a inclus suplimentar, la solicitarea Comisiei Europene, date despre organizații de cercetare, echipe și cercetători. Această versiune este prezentată în detaliu în documentația [11]. CERIF 2000 este format din trei componente:

1. un model complet de date CRIS, care acoperă majoritatea structurilor de date de date a CRIS existente;
2. un set de modele de date, care să constituie studii de caz sau exemple pentru schimbul de date în care sunt implicate sisteme CRIS;
3. un model de metadate, care să furnizeze informație de sinteză cu privire la sursele eterogene de informație.

Modelul de schimb de date și modelul de metadate sunt de fapt subseturi ale modelului complet de date CRIS. La rândul său, modelul de metadate este un subset al modelului complet și al modelului de schimb de date. Aceasta înseamnă că orice informație din modelul de schimb de date se găsește în modelul complet. În același timp, orice informație din modelul de metadate se găsește în modelul de schimb de date și în modelul complet. Detalii privind aceste trei componente sunt prezentate în secțiunea 4.2.

La sfârșitul anului 2006, euroCRIS a anunțat versiunea CERIF 2006, a cărei documentație este în curs de definitivare.

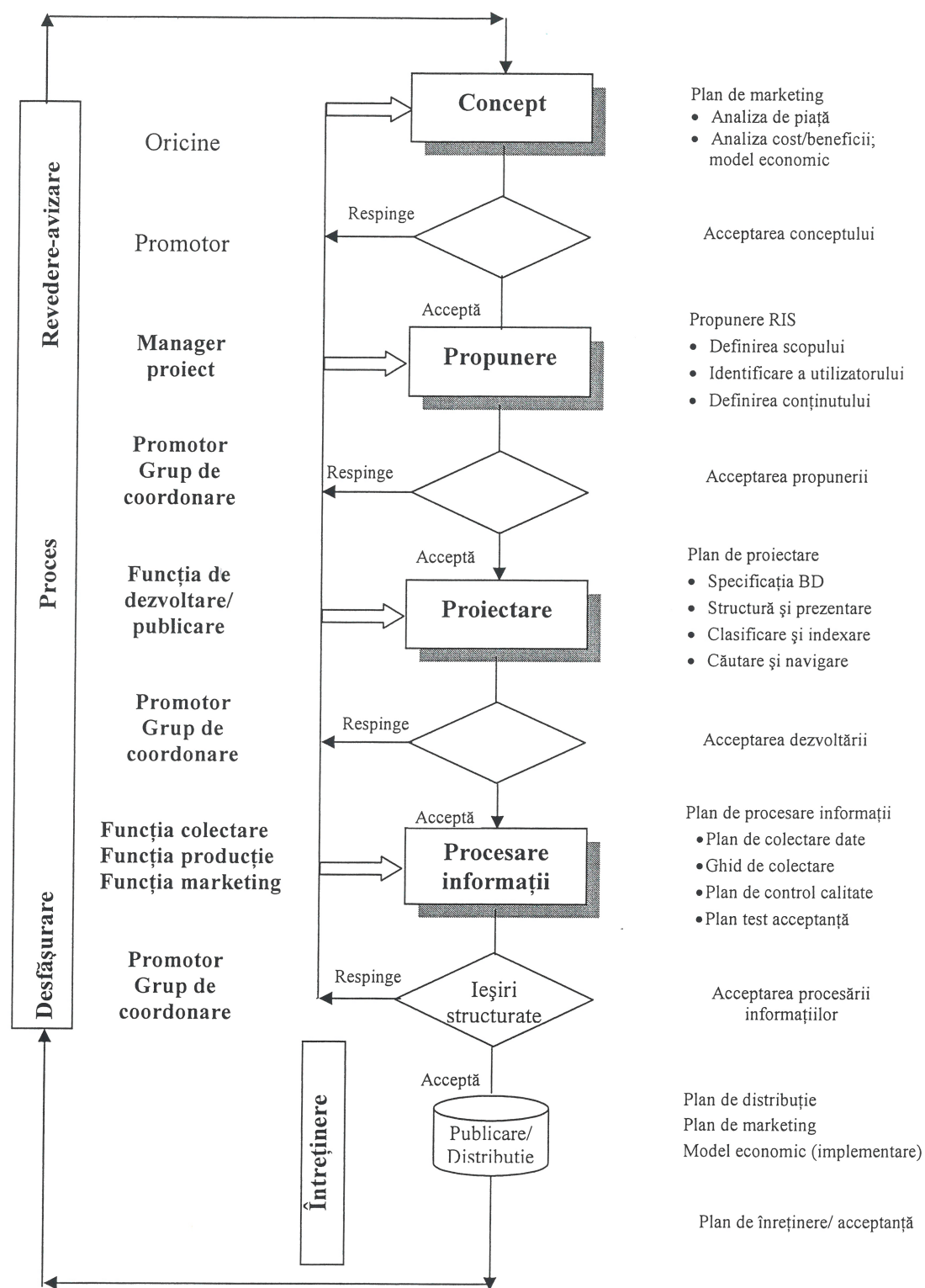


Figura 2. Ciclul de viață CRIS

4.2. Componentele CERIF 2000

Modelul complet de date se bazează pe trei entități principale: Proiect, Persoană și Organizație (figura 3). Relațiile între aceste entități sunt explicitate prin rol și intervalul de timp. Un proiect poate avea relație cu mai multe persoane sau organizații, care pot fi diferențiate prin rolul acestora în cadrul proiectului. În mod similar, o persoană poate participa la mai multe proiecte și poate face parte din mai multe organizații, având în fiecare caz un anumit rol. Fiecare relație de acest gen poate fi caracterizată printr-un interval de valabilitate.

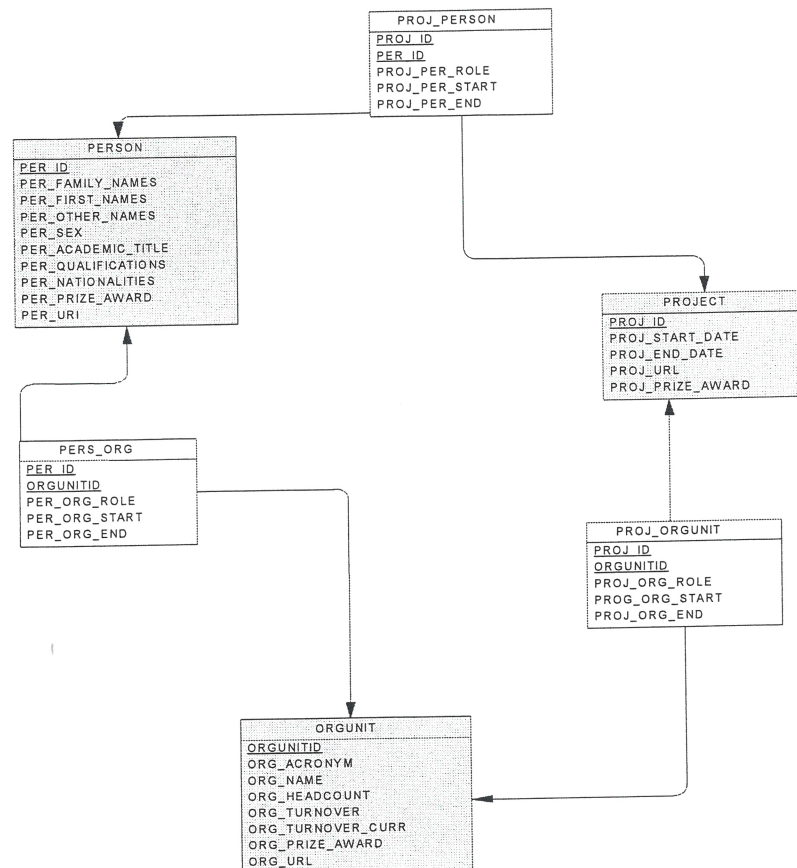


Figura 3. Modelul complet de date CRIS: entitățile de bază și relațiile între ele

În afara acestor trei entități de bază, modelul mai conține:

- entități care descriu rezultatele activității de cercetare (de tip publicații, patente, produse) care sunt asociate celor trei entități principale;
- entități referitoare la echipamente și instalații pentru activitatea de cercetare, asociate cu entitățile Proiect și Organizație. entitatea de tip Serviciu asociată cu entitatea Organizație;
- entitățile de tip Expertiză și CV asociate cu entitatea Persoană;
- entitatea Contact (care furnizează informații de detaliu privind localizarea și posibilitatea de contactare) asociată și entitățile Proiect, Persoană și Organizație.

Modelul complet de date este structurat pe *cinci niveluri de abstractizare*, pentru a facilita înțelegerea lui de către utilizatori. Fiecare nivel reprezintă o detaliere a modelului, după cum urmează:

- nivelul 1 - acoperă cele trei entități de bază menționate anterior;
- nivelul 2 - include entități secundare, ca de exemplu, Rezultate, Clasificare, Contact, Eveniment;
- nivelul 3 - include traduceri în alte limbi ale tuturor entităților de pe nivelul 2;
- nivelul 4 - include entități de tip Nomenclator, care cuprind seturi de valori acceptate pentru diverse atribute;
- nivelul 5 - descrie relațiile mulți-la-mulți între diversele entități.

Modelul de schimb de date se referă generic la varietatea de structuri destinate schimbului de date între diverse sisteme CRIS și aplicații informatice ale utilizatorii finali. Majoritatea acestor modele au ca entitate comună informația de Contact și cel puțin una din cele trei entități de bază. Alte grupe de entități și relații utilizate în aceste modele sunt următoarele:

- Expertiză (competență, CV);
- Rezultat (patente, produse, publicații);
- Instalație și Echipament (servicii, instalații de uz general, echipamente specializate)
- Finanțare (programe de finanțare).

Pot exista anumite informații dintr-o bază de date CRIS, care nu sunt cuprinse în modelul complet de date, aceasta având drept consecință imposibilitatea utilizării acestor informații în schimbul de date ale acestui sistem CRIS.

Modelul de metadate are la bază ideea selectării unui set minimal de entități, atribute și relații necesare pentru a localiza cât mai precis informația solicitată de utilizator. Această selectare se bazează pe o reprezentare sintetică a tuturor instanțelor din modelul complet CRIS, care are semnificația de metadate. Metadatele sunt utilizate ca un catalog de instanțe, pe ansamblul unor baze de date CRIS distribuite și eterogene.

4.3. Dezvoltările versiunii CERIF 2006

Din informațiile disponibile la acest moment, principalele noutăți ale acestei versiuni se referă la modificarea concepției de structurare a ansamblului de entități ale modelului complet de date CRIS, precum și la implementarea unui format de schimb de date conceput să asigure consistența și calitatea datelor pe ansamblul aplicațiilor și furnizorilor de date din comunitatea CRIS [12]

4.3.1. Grupe de entități

Modelul complet de date din versiunea CERIF 2006 definește următoarele grupe de entități CERIF:

- entități de bază;
- entități de nivel doi;
- entități de legătură;
- entități dependente de limbă;
- entități de clasificare.

Prima grupă include entitățile Project, OrganizationUnit, Person, ResultPublication. Față de versiunea precedentă, entitatea referitoare la rezultatul de tip publicație al unui proiect a fost inclusă în grupa entităților de bază.

Entitățile de nivel 2 descriu contextul în care sunt plasate entitățile de bază și interacțiunile cu acesta: program de finanțare, CV, aptitudini, echipament, instalație, eveniment, rezultat de tip produs, rezultat de tip patent, serviciu.

Entitățile de legătură implementează conexiunile între entitățile primelor două grupe: între două entități de bază sau o entitate de bază și o entitate de nivel 2.

Entități dependente de limbă rezolvă cerința de multilingvism formulată pentru sistemele de tip CRIS. Instanțele acestor entități reprezintă versiuni în diverse limbi pentru: denumire, cuvinte cheie, activitate de cercetare – în cazul organizației; interese de cercetare și cuvinte cheie – în cazul persoanei; titlu, cuvinte cheie și abstract - în cazul proiectului; titlu, cuvinte cheie și abstract - în cazul publicației.

Entitățile de clasificare sunt entități de tip nomenclator, destinate definirii tipologiei entităților de bază sau de nivel 2, precum și a rolurilor în cazul entităților de legătură. Din acest motiv, grupa entităților de clasificare mai este numită și nivelul semantic al formatului CERIF.

4.3.2. Formatul de schimb de date CERIF 2006

Acest format este implementat prin fișiere XML, conforme cu recomandarea W3C, versiunea august 2006. Avantajul fișierelor CERIF XML constă în faptul că permit înțelegerea structurii modelului, fără a necesita cunoștințe privind bazele de date și script-uri SQL.

Formatul CERIF XML este necesar pentru operațiile de schimb de date la nivel fizic, motiv pentru care este conform cu convențiile de nume ale modelului fizic CERIF 2006. Specificațiile CERIF XML includ următoarele componente:

- documentația „CERIF 2006 XML – Data Exchange Format Specification”;
- exemple de fișiere pentru fiecare entitate;
- fișiere tip schemă pentru validare;
- set de exemple pentru roluri și scheme de clasificare (de ex. „Person_OrganisationRoles”, „Person_ProjectRoles”) extrase din studii de caz privind implementarea acestui model.

Documentația formatului de schimb de date detaliază structura fișierelor CERIF XML, precum și procedurile de export și import al datelor specifice acestui format de schimb de date.

5. Concluzii

Definirea și specificarea conceptelor CRIS și CERIF reprezintă rezultate semnificative ale preocupării Comisiei Europene pentru asigurarea unor soluții coerente și eficiente de administrare a informației relevante privind activitatea de cercetare. Preluarea și implementarea acestor specificații la nivel instituțional sau național capătă o importanță deosebită în contextul eforturilor de integrare în Aria europeană de cercetare și de activizare a colaborării internaționale în acest domeniu.

Există deja realizări de referință în acest sens, ca de exemplu sistemul *SICRIS - Slovenian Current Research Information System*, dezvoltat și este întreținut de Institutul de știința informației (IZUM) din Maribor, sub autoritatea și în colaborare cu Agenția slovenă pentru cercetare SRA [13]. Sistemul a fost lansat în 1998 și este structurat pe trei secțiuni care se referă la cercetători, grupuri de cercetare (compuse din personal de cercetare și personal tehnic asociat) și, respectiv, organizații de cercetare, implicate în proiecte finanțate de SRA. În prezent, sistemul cuprinde informații despre aprox. 3.200 de proiecte, aprox. 1.000 de grupuri de cercetare și 11.000 de cercetători, peste 600 de organizații de cercetare.

În contextul acestor preocupări pe plan european, ICI București a inițiat un proiect de experimentare a formatului CERIF 2006 la nivel instituțional. Ca membru euroCRIS, institutul are posibilitatea unui schimb activ de informații privind utilizarea și experiența de implementare a acestui format. Pe baza rezultatelor experimentării, urmează să fie formulată o propunere de dezvoltare a unui sistem CRIS la nivel național.

Bibliografie

1. **UNESCO**. Surveying the National Scientific and Technological Potential, Including the Collection and Processing of Data Management, of the R&D System, Paris, 1970, 251p.
2. **UNISIST**. Study Report on the Feasibility of a World Science Information System, Paris, 1971, 161p.
3. **KREYSA, F.**: SSIE – An Information Centre which Stores Foresight, Journal of Chemical Documentation, 1972.
4. **UNISIST**. Guidelines on the Conduct of a National Inventory of Current Research and Development Projects. UNISIST (UNESCO)/SC/75/WS/13. – Paris, March 1975.
5. **UNESCO**. Information Services on Research in Progress: a World-Wide Inventory. Edited by the Smithsonian Science Information Exchange for the General Information Programme and UNISIST (UNESCO). – Paris, UNESCO, 1982, 220p.
6. **EC**. Recommendation Concerning the Harmonisation within the Community of Research and Technological Development Databases (6 may 1991); Official Journal of the European Communities, OJ L 198, the 13th of July, 1991 (<ftp://ftp.cordis.europa.eu/pub/cerif/docs/cerif1991.htm>).
7. **FINCH, P.** (European Commission, DG XIII-D.2). European Research Gateway On/line: ERGO pilot project. CRIS 98 Conference “CRIS – The Way to Innovation”, 12-14 March, 1998, Luxembourg.
8. **VAN WOENSEL, L.**: (CORDIS, European Commission, DG XIII-D.4). User Needs for Research Information. CRIS 98 Conference “CRIS – The Way to Innovation”, 12-14 March 1998, Luxembourg.
9. **JEFFERY, K.**: (CLRC Rutherford Appleton Laboratory). The Future of CRIS. CRIS 98 Conference “CRIS – The Way to Innovation”, 12-14 March, 1998, Luxembourg.
10. * * *: Comisia Europeană. Code of Good Practice, ianuarie 1998.
11. **CE**. CERIF 2000 Guidelines – Final Report of the CERIF Revision Working Group co-ordinated by DG XIII-D.4, European Commission.
12. **JÖRG, B., O. KRIST, K. JEFFERY, G. VAN GROOTEL**: CERIF2006 XML – 1.1 Data Exchange Format Specification. CERIF 2006 – 1.1 FDM, euroCRIS, iulie 2007.
13. * * *: Slovenian Current Research Information System (<http://sicris.izum.si>).