

INTEGRAREA INFORMAȚIILOR ÎN MEDICINA TRANSLAȚIONALĂ

Adriana ALEXANDRU

adriana@ici.ro

Dora COARDOȘ

coardos@ici.ro

Institutul Național de Cercetare – Dezvoltare în Informatică - ICI București

Rezumat: Medicina translațională este un subiect de interes în lumea medicală. Medicina translațională este procesul prin care rezultatele cercetării efectuate în laborator sunt folosite direct alături de rezultatele clinice pentru a dezvolta noi modalități de a trata pacienții. Componentele esențiale ale cercetării din medicina translațională sunt: accesul la informații, analiza de date și integrarea cunoștințelor. Oamenii de știință și medicii trebuie să fie capabili să integreze datele obținute cu alte date disponibile, să combine informații de la mai multe surse și propriile rezultate cu cunoștințe anterioare. Bazându-se pe rezultatele astfel obținute pentru fiecare pacient în medicina translațională, medicina personalizată propune îmbunătățirea diagnozei și tratamentului fiecărui pacient în funcție de caracteristicile individuale ale acestuia. Articolul are ca scop prezentarea principalelor concepte legate de: medicina translațională, sursele de date și structurile de date medicale, standardele disponibile, soluțiile de integrare a datelor, tehnologiile reprezentative care stau la baza integrării surselor de date eterogene și instrumentele reprezentative pentru integrarea informației.

Cuvinte cheie: medicina translațională, ontologii, standarde, integrarea datelor, semantic Web.

Abstract: Translational medicine is a hot topic in the medical world. Translational medicine is the process by which research results obtained in the laboratory are used together with clinical results to develop new ways to treat patients. Essential components of translational research in medicine are: access to information, data analysis and integration of knowledge. Scientists and doctors must be able to integrate data from other data available to combine information from multiple sources and own results with prior knowledge. Thus, based on the results obtained for each patient in translational medicine, personalized medicine proposes the improvement of diagnosis and treatment of each patient depending on its individual characteristics. This article aims at presenting the main concepts related to: translational medicine, data sources and data structures, medical standards, available data integration solutions, representative technologies for the integration of heterogeneous data sources and tools for information integration.

Keywords: translational medicine, ontologies, standards, data integration, semantic Web.

1. Introducere

Medicina translațională are ca scop integrarea datelor din diferite domenii științifice (de exemplu, biochimie, biologie moleculară, medicină clinică, statistică, epidemiologie, etc). Îmbunătățirea legăturii dintre observațiile clinice și cercetările de bază și cele clinice, nu doar că accelerează procesul de tranziție de la laborator la practica clinică, dar totodată ajută și la înțelegerea unor corelații complexe unde sunt implicați diferiți factori și cunoștințe din diverse domenii medicale.

Datorită eterogeneității, volatilității și autonomiei surselor de date, principalele aspecte abordate sunt legate de *integrarea multiplelor surse de date distribuite* (cum ar fi baze de date publice, terminologii, ontologii și publicații științifice) și de prelucrarea eficientă a informațiilor. În plus, bazele de date moderne includ milioane de variabile.

Cantitatea imensă de date care rezultă din progresul științelor biomedicale și dezvoltarea societății informaționale, care are drept scop informatizarea celor mai importante aspecte

ale vieții, în special a celor de sănătate, conduce la utilizarea ontologiilor în sistemele de sănătate și în cercetările biologice.

Un alt aspect de care trebuie ținut cont este *integrarea informației* din punctul de vedere al structurii datelor (de exemplu, formatul în care datele au fost stocate, limbajul folosit, etc.).

În cadrul medicinei translaționale, extragerea informațiilor relevante dintr-una sau mai multe baze de date este o cerință de bază pentru majoritatea sistemelor informatice. În general, aceste sisteme sunt activate de către utilizatorii finali prin transmiterea unor cereri de interogare. Datorită volumului mare de date, se pot considera analogii și metode similare cu cele folosite în cadrul *Semantic Web*.

Medicina personalizată își propune îmbunătățirea procesului de diagnoză și tratament, prin folosirea rezultatelor medicinei translaționale și propunerea unor metode de tratament și medicamente specifice nevoilor fiecărui pacient. Astfel, se dorește eliminarea abordării în care un anumit tratament sau medicament se aplică pe scară largă tuturor pacienților, fără a se ține cont de istoricul

medical sau de specificul fiecăruia. La o analiză mai atentă observăm că ne îndreptăm spre un sistem de terapie personalizată.

Din 2007, Uniunea Europeană a angajat peste 1 miliard Euro în Programul Cadru 7 pentru cercetare, dezvoltare tehnologică și activități demonstrative în scopul finanțării cercetării pentru sănătate, care stă la baza dezvoltării medicinei personalizate. Mergând mai departe, finanțarea în cadrul Orizont 2020 va continua să sprijine acest domeniu.

Esențială pentru realizarea medicinei personalizate este dezvoltarea sistemelor de informare capabile să furnizeze informații exacte și în timp util cu privire la relațiile potențial complexe între pacienți, medicamente și opțiuni terapeutice adaptate.

Cerințele din medicina personalizată includ integrarea cunoștințelor din arhive de date care au fost dezvoltate pentru utilizări diverse și în mod normal nu aderă la o schemă unificată. Acest lucru necesită tot mai mult disponibilitatea datelor și a sistemelor inteligente.

Pornind de la preocupările Uniunii Europene în domeniul cercetărilor pentru sănătate, articolul propune analiza informațiilor legate de stadiul actual în domeniu și anume a soluțiilor de integrare a datelor în medicina translațională prin furnizarea de informații asupra surselor și categoriilor de date, a standardelor disponibile și a soluțiilor de integrare a datelor (Capitolul 2). De asemenea sintetizează modul de utilizare în medicina translațională a ontologiilor asociate bazelor de cunoștințe (definire ontologii, limbaje informatice pentru reprezentarea ontologiilor, utilizarea tehnicilor specifice web semantic pentru integrarea surselor de date eterogene) - Capitolul 3.

2. Tendințe de cercetare, dezvoltare, inovare în domeniul soluțiilor de integrare a datelor în medicina translațională

2.1 Conceptul medicinei translaționale

Medicina translațională este un proces în care descoperirile științifice sunt transferate din cercetarea de bază în cercetarea clinică (1) și apoi rezultatele studiilor clinice sunt

aplicate în practică pentru îmbunătățirea stării de sănătate (2).

Dintre direcțiile de dezvoltare și funcționare ale medicinei translaționale fac parte următoarele:

- dezvoltarea dinamică și aprofundată a biologiei moleculare și celulare;
- stabilirea de standarde înalte pentru pregătirea viitorilor clinicieni care vor lucra în acest domeniu;
- înțelegerea de către cercetători a necesității a bazei raționale din spatele diagnosticului și terapiei;
- stabilirea unei medicine personalizate, bazată pe înțelegerea condițiilor necesare pentru utilizarea unei terapii moleculare noi și personalizate pentru pacienți;
- cooperarea interdisciplinară și transferul cunoștințelor științifice în practica medicală.

Medicina translațională își propune integrarea multi-direcțională a cercetării fundamentale, cercetarea orientată către pacient și către populație cu obiectivul mai general și pe termen lung, îmbunătățirea stării de sănătate creând o punte de legătură dintre cercetările științei de bază și practica clinică. Aceasta presupune o mai bună înțelegere a problematicii sănătății, dezvoltarea de noi standarde, implementarea celor mai bune practici, promovarea interacțiunii între cercetarea de laborator și aplicarea practică.

Medicina translațională reprezintă un domeniu de cercetare multidisciplinar în care se utilizează atât date publice (ghiduri medicale, literatură științifică, baze de date biomedicale), cât și date protejate (date private de la pacienți, date proprietare de la companii farmaceutice și de publicare) [1]. Produsele cercetării translaționale, cum sunt testele de diagnosticare moleculară, sunt principalii factori care potențiază medicina translațională.

Medicina translațională se referă la procesul prin care rezultatele cercetărilor efectuate în laborator sunt folosite direct pentru a dezvolta noi metode de tratare a pacienților. Depinde de integrarea cuprinzătoare a tuturor datelor pacientului cu datele colectate în clinică pentru a facilita și a evalua dezvoltarea medicamentelor [2].

Medicina translațională a fost o prioritate pentru programele naționale de cercetare biomedicală din întreaga lume. Premiile NIH Clinical and Translational Science Awards (CTSAs), prezentate de Zerhouni [3], asigură un punct de referință în cercetarea translațională și s-au dovedit fructuoase în realizarea de proiecte informatice translaționale semantice [4].

În Europa, Kamel și colaboratorii [5] au introdus Innovative Medicine Initiative (IMI), o acțiune comună între Uniunea Europeană și Federația Europeană a Industriilor și Asociațiilor Farmaceutice (EFPIA).

Informatica utilizată în medicina translațională a fost mult timp un caz de utilizare pentru semantica biomedicală. Lucrările de început ale HCLSIG (Health Care and Life Sciences Interest Group) au arătat potențialul tehnologiilor Web-ului semantic pentru cercetarea translațională [6].

2.2 Surse de date, structuri de date medicale

Privind Sistemul Informatic Medical, Organizația Mondială a Sănătății propune combinarea datelor obținute din fișele electronice ale pacienților cu datele de monitorizare de la dispozitivele medicale, de la sistemele de tratament, cu rezultatele laboratoarelor automatizate, arhivele de imagini medicale, informația financiară și mijloacele moderne de schimb de informații (poșta electronică, Internet, videoconferințe) etc.

Date medicale – sunt colectate folosind metode de colectare și de măsurare diferite (cum ar fi examinări fizice, teste de laborator, teste de radiologie, interviuri cu pacientul). Având în vedere că echipamentele de măsurare au diferite intervale de precizie iar medicii au niveluri diferite de educație și de experiență atunci valorile măsurate pot fi dificil de comparat între pacienți. Pe de altă parte, introducerea datelor pacientului în EPR-uri (Electronic Patient Records - Fișa electronică a pacientului) ar putea afecta, acoperirea și coerența surselor de date. Această problemă poate fi rezolvată prin utilizarea unui software mai bun, care să permită introducerea de date structurate eficiente și adecvate și un control mai strict al calității.

Date de biologie moleculară - sunt

volum de date de dimensiuni foarte mari obținute prin utilizarea de noi metode analitice utilizate în studii genomice, de transcripție, proteice și metabolice [7], etc și stocate în baze de date sau de cunoștințe disponibile în toată lumea. Pentru cercetarea viitoare sunt considerate cruciale reprezentarea datelor, adnotarea genelor în baze de date mari și etichetarea lucrărilor de cercetare cu concepte din vocabulare controlate.

Date sensibile pentru pacient - sunt informații cu caracter sensibil și sunt, prin urmare, protejate de legile privind confidențialitatea. Odată cu creșterea mobilității cetățenilor în Uniunea Europeană, schimbul de istoric medical al pacienților devine o necesitate. Cu toate acestea, legislația referitoare la manipularea datelor medicale face dificilă sau chiar imposibilă obținerea permisiunilor necesare pentru utilizarea datelor în afara granițelor țării.

Multe sisteme informatice de spital au fost construite în mod tradițional, neputându-se asigura interoperabilitatea cu alte aplicații în domeniul medical. Schimbul de date ale pacientului cu alți factori interesați din afara organizației este astfel limitat, deși un astfel de schimb poate fi motivat atât din punct de vedere etic, cât și medical.

Există preocupări etice similare pentru datele de biologie moleculară, deoarece datele genetice de la pacienți pot fi folosite pentru a identifica o persoană.

În 2008, Claerhout et al. [8], descrie arhitectura de protecție a datelor în Uniunea Europeană realizată în proiectul Avansarea studiile clinico genomice (ACGT). Obiectivul de nivel înalt al proiectului ACGT era dezvoltarea de metode și sisteme de software care pot îmbunătăți înțelegerea datelor de cercetare a cancerului prin integrarea și analiza informațiilor biomedicale.

2.3 Categoriile de informații în medicina translațională

Domeniul medicinei translaționale face apel la folosirea unor multiple surse de informații eterogene care alcătuiesc bazele de date și cunoștințe medicale.

Descrierea cunoștințelor dintr-un anumit domeniu poate fi realizată doar prin construcții cu o mare varietate de componente

și relații între aceste componente.

Funcționarea corectă a unui sistem decizional medical bazat pe cunoștințe depinde de modul în care se realizează baza de cunoștințe.

Într-o bază de cunoștințe medicale, analiza cunoștințelor este posibilă odată cu specificarea declarativă a termenilor dați. Analiza formală a termenilor este extrem de valoroasă atunci când se încearcă refolosirea ontologiilor existente, dar și când se vrea extinderea lor.

Dintre scopurile pentru care, în gestiunea cunoștințelor, se construiesc *ontologii* menționăm necesitatea definirii unor structuri standardizate pentru informațiile stocate în formatele inițiale diferite în care au fost colectate și a găsirii unor elemente de legătură între acestea, cât și nevoia de a exista o înțelegere comună asupra informației pentru toți cei implicați, oameni sau agenți software.

Pornind de la categoriile de informații definite, un sistem informatic aferent medici+nei translaționale include multiple surse de informații eterogene de volum deosebit de mare, de tip *Big Data*, care sunt distribuite geografic. Mai mult, nu este suficient să se considere că sistemul conține numai date stocate în baze de date, trebuie luate în seamă și cunoștințele domeniului stocate în baza de cunoștințe. Singura soluție pentru considerarea bazelor de date și bazelor de cunoștințe distribuite geografic ca asigurând cu informații un sistem integrat este ca acestea să fie federalizate. Atât sistemele de gestiune ale bazelor de date cât și sistemele expert pot asigura tehnic integrarea într-un sistem federalizat.

2.4 Standarde disponibile

Standardizarea în domeniul medical devine tot mai importantă datorită utilizării frecvente a fișei electronice a pacientului. Standardele facilitează comunicarea automată între diverse sisteme, furnizează informații utile în practica medicală.

Tendențele la nivelul Comunității Europene sunt: integrarea sistemelor medicale și asigurarea interoperabilității.

În Europa, organismul recunoscut în domeniul standardelor pentru informatica medicală este CEN (Comité Européen de

Normalisation - Comitetul European de Standardizare). Acesta are competența și responsabilitatea organizării, coordonării și monitorizării activității de dezvoltare a standardelor în domeniul informaticii medicale, precum și a promulgării acestor standarde.

La ora actuală există în întreaga lume o mare preocupare pentru dezvoltarea standardelor în domeniul sistemelor informatice medicale în general dar și standarde specifice domeniului EHR (Electronic Health Records - Dosarul electronic de sănătate). Aceste standarde acoperă unele aspecte mai generale, cum sunt definirea și scopul sistemelor EHR, tipuri de date și sisteme de codificare a informației, mergând până la specificarea în detaliu a unor modele informatice pentru stocarea datelor ca și a documentelor și modalităților pentru schimbul de informații medicale [9].

Adoptarea pe scară largă a înregistrărilor personale de sănătate depinde de numeroși factori: obstacole în sistemul de sănătate, probleme legate de încredere și confidențialitate, lipsa standardelor tehnice de interoperabilitate și infrastructurii tehnologiilor informaționale de sănătate, divizarea digitală, realizarea incertă cost-eficiență și incertitudinea cererii de pe piață. Există totuși standarde pentru înregistrările personale de sănătate.

Principalele standarde relevante pentru dezvoltarea sistemelor de *evidență electronică a datelor medicale* sunt : ISO DTR 20514 - EHR definition and scope; ISO TS 18308 - EHR Requirements; CEN TS 14796 - Data Types; CEN / TC 251 EN 13606 - EHR Communications; HL7 - EHR Functional Specification; HL7 - Templates specification; HL7 - Clinical Document Architecture; DICOM - Digital Imaging and Communications in Medicine; EDIFACT, XML - Messaging standards.

Multe state europene, inclusiv România, s-au orientat către aceste standarde, majoritatea de origine nord americană, reprezentative fiind HL7 pentru mesaje medicale și DICOM pentru imagistica medicală.

HL7 s-a impus ca standard pentru schimbul de informație pe suport electronic în domeniul sănătății, atât în domeniul clinic cât și în cel administrativ. Obiectivul major al

standardului HL7 îl constituie facilitarea schimbului de mesaje între aplicațiile ce gestionează date medicale;

DICOM - acronim pentru Digital Imaging and Communication in Medicine – este standardul industrial pentru transferul imaginilor medicale și pentru comunicația dintre aparatele electronice medicale. Toate sistemele de stocare și transmitere a imaginilor medicale trebuie să fie conforme cu standardul DICOM, lucru care, la ora actuală, rămâne doar un deziderat, ca urmare a costurilor mari implicate.

2.5 Soluții de integrare a surselor de informații

În *medicină*, ca de altfel în multe alte domenii ale cunoașterii și practicii umane, referitor la dezvoltarea aplicațiilor informatice moderne, problema principală constă în integrarea diverselor surse de date și cunoștințe eterogene existente. Integrarea informațiilor se confruntă cu următoarele probleme:

- *eterogenitatea structurală* - se referă la faptul că diferite sisteme de aplicații stochează datele proprii în structuri diferite;
- *eterogenitatea semantică* - se referă la înțelesul diferit pe care îl pot avea elementele, ce conțin o informație, în sistemele de aplicații;
- *problema inconsistenței și a redundanței datelor* în diferite sisteme de aplicații.

Există, desigur, *multe abordări de integrare a datelor*, dar subliniem doar *două abordări* care au fost aplicate anterior în domeniul informaticii medicale: *depozite de date* (data warehousing) și *abordări bazate pe mediere*.

Depozitul de date medicale reprezintă o colecție de date integrată, variabilă în timp, non-volatilă, orientată către subiect care este utilizată în sprijinul luării deciziilor de management. Datele de la sursele de date sunt încărcate, transformate, curățate și actualizate în mod regulat. Pentru analize multidimensionale sunt utilizate instrumentele de prelucrare analitică on-line (OLAP).

Stabilirea unor depozite de date nu este întotdeauna o soluție optimă. De exemplu, datele de dimensiuni enorme, fiind obținute din proiectele de cercetare în domeniul genomului uman, nu sunt transferate cu ușurință.

Prin urmare, în loc de colectarea și stocarea datelor într-un depozit centralizat, o abordare alternativă este lăsarea datelor la sursele de date originale și combinarea lor în mod dinamic prin interogări rezolvate de un mediator.

În procesul de integrare a datelor, un *sistem bazat pe mediator* realizează translația dinamică a interogărilor utilizatorului între schema unei componente mediator și schemele surselor de date. Un astfel de sistem este ilustrat în Figura 1, în

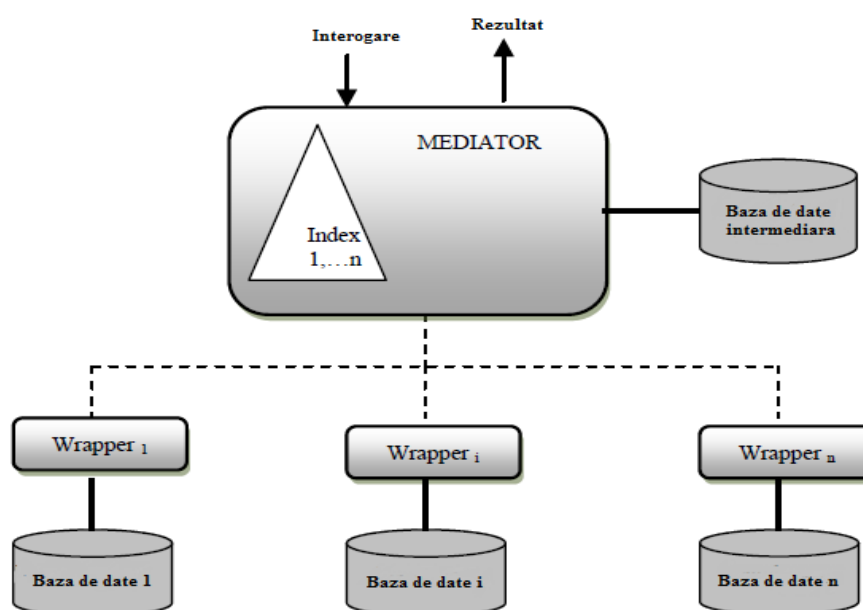


Figura 1. Arhitectura unui sistem bazat pe mediator [11]

care wrapper-ul este un set de programe care accesează sursele de date și asigură vizualizarea datelor într-o schemă de date uniformă.

În principiu, integrarea datelor cu cunoștințele, într-un sistem informatic medical care se bazează în bună parte pe IA, se realizează pe două căi:

- construirea ontologiilor și utilizarea tehnologiilor specifice web-ului semantic ca primă etapă a proiectării care să permită integrarea bazată pe semantica datelor și cunoștințelor [10];
- realizarea unor interfețe între bazele de date cu informație medicală, respectiv între acestea și bazele de fapte ale sistemelor expert corespunzătoare diverselor domenii ale medicinei.

Integrarea datelor în cadrul medicinei translaționale și utilizarea cunoștințelor incluse în cadrul ontologiilor sunt premise pentru îmbunătățirea cunoașterii bolilor și pentru îmbunătățirea proceselor de diagnosticare și de prognoșticare a evoluției acestora.

Pentru atingerea acestor obiective de integrare sunt necesare:

(a) crearea unui model semantic care să reprezinte surse heterogene de date și integrarea acestora utilizând tehnologii Web semantice și ontologiile existente;

(b) dezvoltarea unei metodologii care să exploateze cunoștințele existente în cadrul ontologiilor și să îmbunătățească rezultatele metodelor de explorare a cunoștințelor obținute cu seturile de date ale medicinei translaționale [1].

Integrarea tuturor acestor date este o problemă ce se pune în mod stringent.

3. Utilizarea de ontologii asociate bazelor de cunoștințe în medicina translațională

3.1 Utilizarea Ontologiilor în medicina translațională

În domeniul calculatoarelor, ontologia este considerată o modalitate *de definire și specificare a obiectelor, conceptelor, entităților* din lumea reală, precum și a *relațiilor* dintre ele.

Entitățile caracteristice ontologiilor sunt: *clase, obiecte, attribute, funcții, restricții, reguli, axiome și evenimente*, manipulate prin limbaje algebrice sau pictoriale. Deoarece în medicina translațională volumul de informații este foarte mare și sursele lor sunt foarte diferite nu se poate începe procesul de analiză și structurare de la *clase*.

Este necesar lucrul cu un concept de rang mai înalt decât *clasa*, este vorba de *categorie*. Categoriile vor fi definite de nivelurile semantice pe care specialiștii domeniului le consideră mai importante. Astfel acestea se raportează la: organizarea medicală existentă, individ, activități derulate de cadrele medicale, asistenții medicali.

Practic, ontologia este un vocabular și o specificare a unei colecții de concepte, a simbolurilor lor și a relațiilor dintre ele (inclusiv fapte implicate), precum și o caracterizare a semnificației acestora.

Obiectivul unei ontologii constă în reprezentarea unei conceptualizări, partajabile și reutilizabile, în care sunt ignorate detaliile specifice aplicațiilor.

Sfera de cuprindere a unei ontologii se referă la toate aplicațiile domeniului, nu numai la o anumită aplicație.

Activitatea de construire a unei ontologii este complexă și presupune cercetarea intensă a domeniului.

Pentru a utiliza în mod eficient în cadrul aceluiasi sistem surse multiple de date eterogene, se propune folosirea de *ontologii* asociate fiecărei surse de informație, având principalul rol de a exprima prin structuri standardizate informațiile stocate în bazele de cunoștințe.

Ontologiile sunt din ce în ce mai mult folosite în aplicațiile și sistemele care au la bază interogări semantice. În acest context, principalul avantaj al ontologiilor este faptul că, prin intermediul unui vocabular structurat, pot descrie baze de cunoștințe, chiar dacă acestea nu sunt organizate într-o structură de baze de date relaționale. Utilizând ontologii în cadrul proceselor de interogare se pot obține performanțe mai bune și rezultate exacte la interogările utilizatorului. În plus, ontologiile sunt preferate în interogările semantice, fiind mai ușor de integrat în cadrul aplicațiilor de interogare a datelor decât

folosirea directă a informațiilor din cadrul bazelor de date relaționale.

Datorită domeniului medical vast abordat de către medicina translațională, este necesară utilizarea a numeroase ontologii, dintre cele mai complexe sau variate. În plus, ontologiile nu sunt văzute ca fiind sursa unică de date, ci mai degrabă ca fiind componenta care face legătura între utilizator și informațiile concrete stocate în multiple baze de cunoștințe.

Ontologiile au fost dezvoltate pentru reprezentarea entităților de interes biomedical și relațiile dintre ele, în multe domenii și diverse niveluri de granularitate.

Sutele de ontologii biomedicale dezvoltate până în prezent sunt accesate de către comunitatea științifică în principal prin intermediul a două mari resurse: *BioPortal* [12] și *Ontology Look-up System (OLS)* [13].

Ambele resurse sunt foarte diverse din punct de vedere al conținutului și este dificilă caracterizarea lor precisă. Un mare număr de ontologii se regăsesc în ambele resurse.

3.2 Limbaje informatice pentru reprezentarea ontologiilor

În funcție de scopul construirii ontologiei există mai multe metode de construcție și respectiv limbaje.

În informatică și inteligență artificială, limbaje pentru descrierea de ontologii sunt limbaje formale utilizate pentru a construi ontologii. Ele permit codificarea cunoștințelor despre anumite domenii și adesea includ reguli de raționament care susțin procesarea cunoștințelor.

Au fost dezvoltate mai multe limbaje informatice pentru reprezentarea ontologiilor în contextul web-ului semantic.

Principalele limbaje de definire a ontologiilor sunt bazate pe XML - un limbaj care este foarte ușor interpretabil de către calculatoare.

RDF (Resource Description Framework) și RDFS (Resource Description Framework Schema) sunt limbajele standard de reprezentare a datelor în Web-ul semantic.

Limbajul recomandat de marcare ontologii pentru Web-ul semantic în scopul codificării cunoștințelor este OWL (Web

Ontology Language).

Datele pot fi navigate sau interogate cu SPARQL (Simple Protocol and RDF Query Language), maparea dintre RDF și scheme trebuind făcută de fiecare dată când sunt accesate datele.

3.3 Tehnici specifice web semantic pentru integrarea surselor de date eterogene în medicina translațională

Medicina translațională necesită tehnologii adecvate care să permită interpretarea corectă a resurselor de date distribuite, precum și o infrastructură flexibilă care să păstreze semantica datelor la toate nivelurile de integrare.

În domeniul medicinei translaționale, integrarea datelor și interoperabilitatea sunt cerințe importante pentru analiza eficientă a datelor.

Una din soluțiile propuse pentru satisfacerea acestor cerințe constă în utilizarea web-ului semantic și a tehnologiilor acestuia.

Web-ul Semantic permite regăsirea automată a datelor, din domenii ce par fără legătură, pentru a le combina în vederea găsirii de relații și a face descoperiri [14].

Web-ul semantic este capabil de efectuarea operației de luare a deciziilor [15]. Pentru aceasta este necesar ca datele să fie stocate în *mod ontologic*, ceea ce permite o interpretare mai inteligentă a acestora.

RDF este limbajul standard de reprezentare a datelor în Web-ul semantic. Dacă resursele care trebuie integrate sunt în acest format nu este necesară o transformare a acestora. Dacă acestea sunt în alte formate, cum ar fi cel relațional sau XML, este necesară o transformare. O a treia alternativă se aplică în cazul datelor stocate în baze de date relaționale, acestea fiind expuse ca date în RDF utilizând maparea schemei bazei de date la o ontologie în format RDFS sau OWL. Această opțiune permite evitarea replicării datelor.

Dezvoltarea Web-ului Semantic și a limbajelor acestuia de descriere a resurselor (*Resource Description Framework - RDF*), a formatelor de interschimb al datelor (RDF/XML), și a notațiilor, cum este *RDF Schema* (RDFS), și *Web Ontology Language* (OWL),

a permis furnizarea unei descrieri formale a conceptelor, termenilor și a relațiilor dintr-un anumit domeniu de cunoaștere. OWL furnizează definiții computaționale formale și instrumente pentru raționamente, care permit dezvoltarea și

unui pacient, prin care, pornind de la probele recoltate de la pacient, se realizează teste de genomică, acestea fiind utilizate împreună cu observațiile clinice și cu arhivele de date integrate, pentru realizarea unei interpretări corecte a analizelor este prezentat în Figura 2.

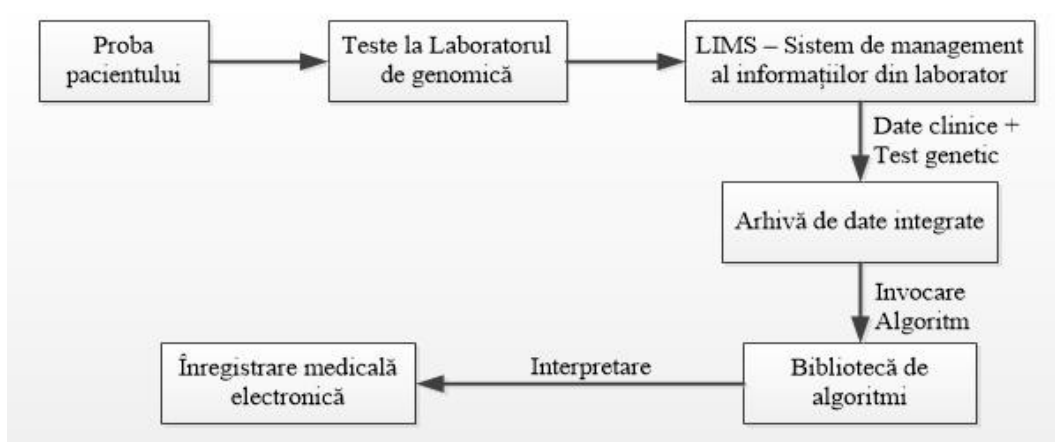


Figura 2. Flux de informații în Medicina Translațională (adaptare după [14])

întreținerea ontologiilor. Multe ontologii din știința sănătății au fost convertite în OWL.

Integrarea datelor obținute de la *Înregistrările medicale electronice* (EMR) cu datele genomice obținute de la laborator LIMS (Laboratory Information Management System - Sistem de management al Informațiilor din Laborator) se poate realiza utilizând specificațiile Web-ului semantic (RDF, OWL). Această abordare cuprinde patru etape:

- crearea unei ontologii a domeniului care identifică cele mai importante concepte ale domeniilor clinic și genomic;
- proiectarea și crearea de wrapper care expun datele dintr-o arhivă într-o vedere RDF;
- specificarea regulilor de mapare care furnizează legături între datele ce sunt regăsite în diferite arhive de date;
- crearea unei interfețe utilizator pentru: (a) specificații de mapare a legăturilor dintre date și (b) vizualizarea informației integrate.

Pentru realizarea procesului de integrare a datelor este nevoie ca utilizatorul final (cercetător genomic sau un cercetător clinician) să interacționeze cu sistemul care afișează grafuri RDF din surse diferite de date.

Fluxul de informații asociat examinării

Concluziile examinării se completează în înregistrarea medicală a pacientului.

4. Concluzii

În cadrul acestui articol a fost făcută *analiza conceptelor, datelor și cunoștințelor aferente medicinei translationale*, un domeniu de mare actualitate în medicină, care utilizează metode din domeniul biologiei sistemelor și tehnologiei informației pentru integrarea cunoștințelor și datelor din mai multe domenii, atât la nivelul individului, cât și al întregii populații.

Gradul ridicat de *eterogenitate a datelor* poate avea un efect negativ asupra rezultatelor finale și duce la creșterea complexității procesului de analiză în cadrul sistemelor informatice aferente.

În contextul medicinei translaționale, se dorește crearea unei *baze de cunoștințe*, formată în principal din informații teoretice obținute prin asocierea mai multor ontologii și baze de cunoștințe, informații clinice obținute din diverse cazuri din practică, statistici care să completeze rezultatele obținute din interogarea datelor teoretice și clinice, cât și orice alte surse de date care pot completa sau valida rezultatele obținute din analiza surselor de date menționate anterior.

Pentru a utiliza în mod eficient în cadrul

aceluiași sistem surse multiple de date eterogene, sunt sintetizate informațiile.

Pentru *integrarea multiplelor surse de date eterogene* din domeniul medicinei translaționale, considerăm că sunt necesare următoarele abordări:

- folosirea de ontologii asociate multiplelor surse de date eterogene care alcătuiesc bazele de cunoștințe medicale, cu rolul de a defini structuri standardizate pentru datele stocate în formatul inițial în care au fost colectate, fără prelucrări anterioare, și de a găsi elemente de legătură între acestea;
- folosirea metodelor aplicate în cadrul Semantic Web ca și componente principale ale unei arhitecturi care să faciliteze obținerea de relații relevante dintre multiple baze de cunoștințe eterogene.

Printre preocupările viitoare ale colectivului elaborator se va include abordarea aspectelor legate de elaborarea unui model arhitectural al unui mediu de integrare a surselor de informații folosite în medicina translațională la nivel semantic și de interogare distribuită a surselor de date biomedicale.

BIBLIOGRAFIE

1. **MACHADO, C. M.:** Knowledge Representation for Data Integration and Exploration in Translational Medicine. Doutoramento em informática especialidade bioinformática, 163 pag., 2013.
2. **WOOLF, S.:** The meaning of translational research and why it matters. JAMA, Vol. 299, Issue 2, pp. 211-213, 2008.
3. **ZERHOUNI, E.:** Translational and clinical science—time for a new vision. New England Journal of Medicine, Vol. 353, Nr. 15, pp. 1621-1623, 2005.
4. **CHEN, H.; YU, T.; CHEN, Y.:** Semantic web meets integrative biology: a survey. Briefings in bioinformatics, Vol. 14, no. 1, pp. 109-125, 2013.
5. **KAMEL, N.; COMPTON, C.; MIDDELVELD, R.; HIGENBOTTAM, T.; DAHLÉN, S.:** The Innovative Medicines Initiative (IMI): a new opportunity for scientific collaboration between academia and industry at the European level. European Respiratory Journal. Vol. 31, Issue 5, pp. 924-926, 2008
6. **RUTTENBERG, A.; CLARK, T.; BUG, W.; SAMWALD, M.; BODENREIDER, O.; CHEN, H.; DOHERTY, D.; FORSBERG, K.; GAO, Y.; KASHYAP, V.; KINOSHITA, J.:** Advancing translational research with the Semantic Web. BMC Bioinformatics, Vol. 8 (Suppl 3), S2, 2007.
7. **BRANESCU, I.; PURCAREA, V. L.; DOBRESCU, R.:** Solutions for medical databases optimal exploitation. Journal of medicine and life, Vol.7, Issue 1, pp. 109-118., 2014 Mar 15.
8. **CLAERHOUT, B.; FORGO, N.; KRUGEL, T.; ARNING, M.; DE MOOR, G.:** A Data Protection Framework for Transeuropean genetic research projects. Studies in health technology and informatics. Vol. 41, pp. 67-73, 2008.
9. **TAO, C.; JIANG, G.; WEI, W.; SOLBRIG, H. R.; CHUTE, C. G.:** Towards semantic-web based representation and harmonization of standard meta-data models for clinical studies. AMIA summits on translational science proceedings, pp. 59-63, 2011.
10. **MACHADO, C. M.; REBHOLZ-SCHUHMAN, D.; FREITAS, A. T.; COUTO, F. M.:** The semantic web in translational medicine: current applications and future directions. Briefings in bioinformatics 16, no. 1 (2015): 89-103.
11. **BRANESCU-RASPOP, I.; PURCAREA, V. L.; DOBRESCU, R.:** Improving Query Mechanisms for Biomedical Ontologies, International Journal of Biology and Biomedical Engineering, Volume 9, 2015.
12. <http://bioportal.bioontology.org/>
13. <http://www.ebi.ac.uk/ols/beta/>

14. **SIKOS, L. F.:** Mastering Structured Data on the Semantic Web – From HTML5 Microdata to Linked Open Data-Make your Web Contents Machine-Interpretable using Artificial Intelligence, Apress, ISBN-13 (electronic): 978-1-4842-1049-9, 2015.
15. **JANBANDHU, R. R.; KARHADE, V.; SELOTE, R.:** Ontological Framework for Semantic Web, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 12, 2015.
16. **KASHYAP, V.; HONGSERMEIER T.:** Can Semantic Web technologies enable translational medicine? In: C.J.O. Baker CJO, K.H. Cheung (Eds.). Semantic Web: Revolutionizing knowledge discovery in the life sciences, pp. 249-277, Springer Verlag, 2007.