DIGITIZAREA ȘI PREZENTAREA ALFABETULUI CHIRILIC ROMÂNESC DE TRANZIȚIE (1830 – 1860)

Cercetător științific **Valentina DEMIDOVA**
Doctor în științe informatice **Ludmila BURȚEVA**
Institutul de Matematică și Informatică al AȘM

1830-1860 ROMANIAN CYRILLIC DIGITIZATION AND PRESENTATION

Summary. Presented paper concerns problems which arise together with growing of databases of scanned ancient texts. Historical text saved as image is unsuited for modern methods of textual information processing. Automatic digitization resolves this problem, but requires specific methods for representation of unusual today scripting. This paper presents the solution for digitizing of Romanian Cyrillic of printed texts of transitional period 1830–1860.

Keywords: automatic text digitization, Romanian Cyrillic, decyrillization.

Rezumat. Lucrarea prezentată se referă la problemele care apar odată cu creșterea volumului bazelor de texte vechi scanate. Textul istoric păstrat ca imagine nu este potrivit pentru metodele moderne de procesare a informațiilor textuale. Digitizarea automată rezolvă această problemă, dar necesită metode specifice pentru reprezentarea scriptului neobișnuit la ziua de azi. Această lucrare prezintă o soluție pentru digitizarea textelor tipărite în alfabetul Chirilic Românesc în perioada de tranziție 1830–1860.

Cuvinte-cheie: digitizarea automată a textelor, alfabetul chirilic românesc de tranziție, dechirilizare.

INTRODUCERE

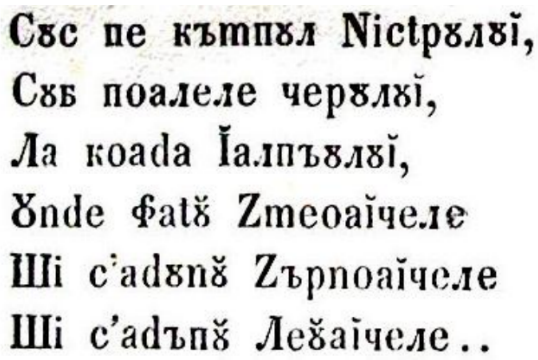
Prin dechirilizare înțelegem procesul de translitera-re a limbii române din scrierea în grafia chirilică din perioada de tranziție (1830–1860) în cea cu grafie latină.

De ce problema dechirilizării a devenit importantă? Cauza principală constă în faptul că cercetările actuale în domeniul conservării patrimoniului cultural s-au concentrat pe particularitățile fiecărei culturi europene [1]. Textele scrise în alfabetul chirilic românesc de tranziție, spre deosebire de cele scrise în grafia chirilică contemporană, sunt răspândite nu doar în Moldova, ci și în România. Citirea acestor texte, care constituie un patrimoniu vast [2], prezintă dificultăți în special pentru specialiștii care nu sunt familiarizați cu grafia chirilică. Odată cu epoca digitală devine oportună și posibilă digitizarea acestor texte vechi, existente doar pe hârtie, astfel ca ele să fie accesibile tuturor.

Drept exemplu, vom apela la un fragment din balada lui Vasile Alecsandri (figura 1), care a fost publicat cu alfabetul chirilic românesc de tranziție, transpus, în paralel, în grafia latină.

1. VIZUALIZAREA SIMBOLURILOR ALFABETULUI CHIRILIC ROMÂNESC DE TRANZIȚIE

Problema vizualizării textelor istorice apare odată cu digitizarea integrală a acestora, în loc de stoca-


Сѹс пе кътпѹл Nістрѹлѹї,
Сѹб поалеле черѹлѹї,
Ла коада Їалпъѹлѹї,
Ѹnde фатѹ Zmeoaїчеле
Ші с'adѹnѹ Zърноаїчеле
Ші с'adъпѹ Леѹаїчеле..

Sus pe cămpul Nistrului,
Sub poalele cerului,
La coada Ialpăului,
Unde fat' Zmeoaicele.
Și s-adun' zărnoaicele
Și s-adăp' Leuaicele..

Figura 1. Fragment din balada *Român Grue-Grozovanu* de V. Alecsandri

re a imaginilor scanate. Prin distribuția suportului standardului de Unicode16 pe o varietate de sisteme de operare au fost alocate coduri pentru orice scriere istorică: cuneiformă, hieroglife, scrisoare Celtic etc. Dispunând de codurile corespunzătoare, problema vizualizării poate fi rezolvată prin intermediul fonturilor specializate.

Pentru alfabetul chirilic românesc de tranziție există deja trei fonturi: Kliment Std, Everson, Roman Cyrillic Std care au diferite imagini, dar care cuprind toate cele 43 de caractere „canonice". Cu toate acestea, în timpul formării versiunii canonice a alfabetului de tranziție au apărut litere care nu sunt incluse în setul final. Astfel, textele electronice obținute sunt incomplete, iar specialiștii în filologie istorică rămân lipsiți de informații foarte necesare.

Problema se poate rezolva pe seama codurilor libere existente în diapazonul Unicode-6 acordat pentru chirilică. În plus, există dificultăți caracteristice pentru alfabetul intermediar, de exemplu, utilizarea caracterelor modificate ale alfabetului latin pentru afișarea caracteristicilor fonetice, în special ĭ, ŭ pentru reprezentarea vocalelor scurte. Evident, o astfel de marcă fonetică prezintă mare interes pentru istorici, de aceea ea trebuie să fie stocată în text. Simbolurile relevante sunt disponibile în versiunea extinsă a Unicode latină, astfel problema se rezolvă prin implicarea codurilor suplimentare.

În această lucrare este prezentat un modul al instrumentarului de digitizare a textului istoric din alfabetul chirilic românesc de tranziție. Acest instrumentar [3] a fost elaborat de grupul de lingvistică computațională al Laboratorului „Sisteme de Programare" de la Institutul de Matematică și Informatică al AȘM. Instrumentarul de digitizare oferă utilizatorului acces prin pagina web și permite alegerea perioadei de timp. În funcție de alegerea utilizatorului, se activează modulul respectiv al instrumentarului.

Modulul prezentat face conversia alfabetului chirilic românesc de tranziție în alfabetul latin pentru perioada 1830–1860.

Instrumentarul este scris în limbajul Java [4], pentru care suportul multiplatformei și, prin urmare, suportul Unicode, sunt proprietăți de bază. În cazul în care fontul corespunzător este înregistrat în sistemul de operare, instrumentele de programare de interfață ale limbajului Java rezolvă problema de vizualizare printr-o simplă adresare la această înregistrare.

În tabelul 1 este prezentată corespondența literelor specifice ale alfabetului de tranziție cu scrierea lor în grafia latină și codul UTF16.

Tabelul 1

Corespunderea literelor specifice și codurile lor UTF16

Ѣ	Ea	0462
ѣ	ea	0463
Ѥ	Ia	0465
ѥ	ia	0464
Ѧ	Â	0466
ѧ	â	0767
Ѩ	Î, Îm, În	A64e
ѩ	î, îm, in	A65f
Ꙋ	U	A64a
ꙋ	u	A64b
К	C, Ch (înainte de e, и)	041a
к	c, ch (înainte de e, и)	043a
Ĭ	Ĭ	012C
ĭ	ĭ	012D
Ъ	Ă	042A
ъ	ă	044A
Щ	Şt	0429
щ	şt	0449
Ц	G, Gh (înainte de e, и, ю)	049F
ц	g, gh (înainte de e, и, ю)	044F

2. DESCRIEREA REGULILOR DE TRANSLITERARE (DECHIRILIZARE)

La elaborarea și descrierea regulilor de transliterare a cuvintelor din limba română, scrise în grafie chirilică a alfabetului de tranziție românesc, într-un echivalent scris în grafia latină, ne-am condus de normele gramaticale în vigoare. Vom identifica două tipuri de reguli: elementare și complexe.

Alfabetul chirilic de tranziție pentru limba română s-a modificat pe parcursul utilizării sale, înregistrând circa 20 de versiuni. Cercetările noastre vizează un număr maximal de litere care constituie 43 de simboluri. În lucrarea prezentă examinăm 36 de litere întâlnite în textele analizate, dintre care 27 sunt cele contemporane, plus nouă litere vechi: Ѣ ѣ, Ѥ ѥ, Ѧ ѧ, Ѩ, ѩ, Ꙋ, ꙋ Ĭ ĭ, Ъ ъ, Щ щ, Ц ц. Menționăm că algoritmul nostru permite adăugarea simplă a literelor care ar putea apărea în procesul viitoarei analize a textelor.

În conformitate cu regulile elementare se efectuează transliterarea a 31 de litere. Pentru simbolurile chirilice, care sunt identice pentru toate variantele alfabetului chirilic, noi folosim regulile descrise în [5]. Regulile elementare sunt cele ce efectuează punerea în corespondență independentă de context a unei litere chirilice cu una sau mai multe litere din grafia latină. Literele prezentate în tabelul 2 sunt traduse în baza regulilor elementare.

Tabelul 2

Regulile elementare de transliterare

Chirilic	а	б	в	д	е	ж	з	и	й	л	м	н	о	п	р	с	т	ф	х	ц	ш	ь	э	ю	Ѣ	ѥ	ѧ	ȣ	ѝ	ъ
Latin	***a***	***b***	***v***	***d***	***e***	***j***	***z***	***i***	***i***	***l***	***m***	***n***	***o***	***p***	***r***	***s***	***t***	***f***	***h***	***ţ***	***ş***	***i***	***ă***	***iu***	***ea***	***ia***	***â***	***u***	***ĭ***	***ă***

Rămân cinci litere care necesită o procesare suplimentară și anume: „г”, „к”, „ч”, „ѧ”, „ц”.

Conversiile acestor litere depind de context, ceea ce vom descrie mai jos. Procesul de dechirilizare îl vom prezenta în tabelele 3-6.

Tabelul 3

Reguli pentru literile *„г”* și *„ц”*

г, ц	**gh**	înainte de: e, и, ю
	g	în restul cazurilor

Exemple: ***цениȣ*** **– gheniu, богат – bogat**

Tabelul 4

Reguli pentru litera *„к”*

к	**ch**	înainte de: e, и, ю
кс	**x**	
к	**c**	în restul cazurilor

Exemple: ***кȣпринде*** **– cuprinde,** ***кѥр*** **– chiar**

Tabelul 5

Reguli pentru litera „ѧ”

ѧ	**î**	înainte de: m, n
	îm	înainte de: b, p
	în	în restul cazurilor

Exemple: **ѧн – în , ѧнчепȣт– început, ѧpune – împune**

Tabelul 6

Reguli pentru litera „ч”

ч	**c**	înainte de: e, и
	ce	înainte de: a
	ci	înainte de: o, y
	ci	la sfârșit de cuvânt
	ci	înainte de consoană

Exemple: причина – pricina, фаче – face, чинститъ – cinstită, нторчї – întorci

3. ANALIZA COMPARATIVĂ A PROCESULUI DE TRANSLITERARE PENTRU DIFERITE PERIOADE DE TIMP

În continuare vom face o analiză comparativă a procesului de dechirilizare pentru literele alfabetului de tranziție (1830–1860) cu cel descris în [6] pentru alfabetul și ortografia din perioada 1945–1989. Algoritmul propus în [5] a stat la baza algoritmului prezent. După cum am menționat anterior, în conformitate cu regulile elementare se efectuează transliterarea a 31 de litere. Pentru unele litere (***„г”, „к”, „ч”, „ц”)*** regulile de transliterare, deși nu sunt elementare, coincid cu cele pentru alfabetul chirilic contemporan [6]. Așadar, existența unui surplus de litere (43) a adăugat doar o singură regulă adițională – cea cu litera ѧ, în schimb a rezolvat mai multe probleme care au apărut la procesul de transliterare a chirilicii contemporane. Vom evidenția cele mai importante probleme:

- Dispare problema ce ține de ortografia nouă, datorită faptului că alfabetul de tranziție conține următoarele simboluri pentru sunetul ***„î”*** în poziții diferite:

Tabelul 7

Literele care corespund sunetului *„î ”*

Ѧ	**Â**
ѧ	**â**
	Î, Îm, În
ѧ	**î, îm, in**

- Se rezolvă problema ce ține de transliterarea literei ***„я”.***

După cum a fost descris în mod detaliat în [5], neam confruntat cu mari dificultăți la procesul de transliterare a literei ***„я”***. Aceste dificultăți au fost legate de faptul că transliterarea literei ***„я”*** este ambiguă și depinde de context. În special, această problemă apare în ocurența lui ***„я”*** în interiorul cuvântului. Regula de transliterare poate fi atât „я” – ***„ia”,*** cât și „я” – ***„ea”.*** Menționăm că stabilirea acestei reguli poate fi efectuată doar prin apelarea la dicționarele externe, de exemplu la DEX.

Așa dar, alfabetul de tranziție rezolvă problema ce ține de litera ***„я”***: fără a o conține, dispune de simboluri specializate pentru diftongi:

Tabelul 8

Simbolurile pentru prezentarea diftongilor *„ia”* și *„ea”*

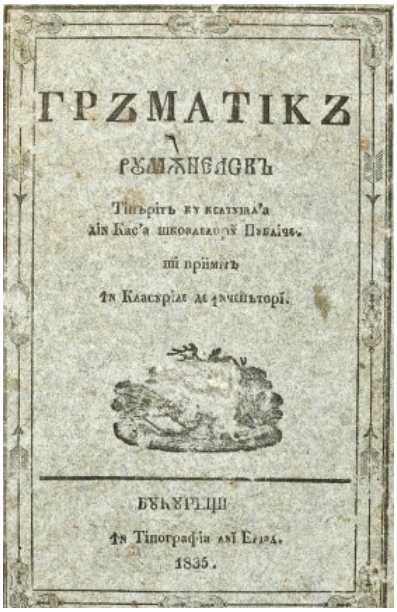
Ѣ	Ea
ѣ	ea
ѧ	Ia
ѥ	ia

Deoarece problemele enumerate mai sus au fost soluționate, transliterarea alfabetului de tranziție a fost efectuată cu succes în 99% de cazuri. Trebuie de menționat faptul că succesul transliterării chirilicii moder-

ne a constituit maximum 91%.

4. EXEMPLE DE TRANSLITERARE

În acest paragraf vom aduce un exemplu de dechirilizare a unui text din ***Gramatica Românească, editată în București, 1835*** (figura 2), care ilustrează complexitatea problemei.


ГРѢМАТІКѢ

РѢМѢНЕСКЪ

Тіпъріть ку кедтуіла'а
ліи Кас'а школелорѝ Пъблічє.

ші пріімітъ
ін Класъріле де ынчіпъторі.

БУКУРЕЩІ
ін Тіпографіа лui Еліад.
1835.

Figura 2. Coperta cărții *Gramatica Românească*

În plus, acest exemplu demonstrează cât de util poate fi instrumentul prezentat pentru specialiști, îndeosebi pentru cei care nu sunt familiarizați în scrisul chirilic:

În figura 3 este prezentat fragmentul textului original.

În continuare urmează textul obținut în urma dechirilizării:

Introducere

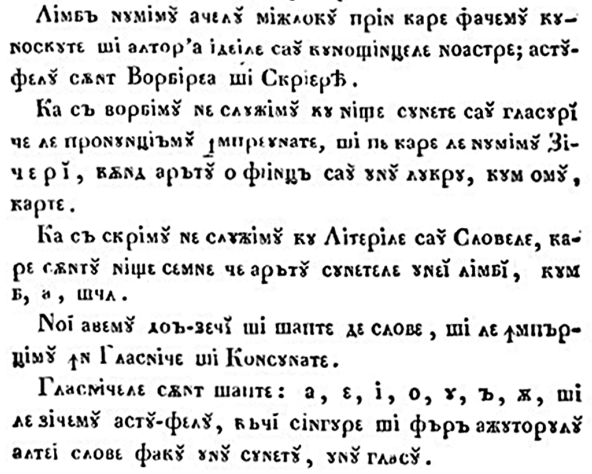
Grămatic'a este unu meșteșugu prin care învățămu a vorbi și a scri o limbă fără greșală.

Limbă numimu acelu mijlocu prin care facemu cu noscute și altor-a ideile sau cunoștințele noastre; astu' felu sânt Vorbirea și Scrierea.

Ca să vorbimu ne slujimu cu niște sunete sau glasuri ce le pronunțiămu împreunate, și pe care le numimu zi'ceri, când arătu o ființă sau unu lucru, cum omu, carte.

Ca să scrimu ne slujimu cu Literile sau Slovele, care sântu niște semne ce arătu sunetele unei limbi, cum b, a, șcil.

Noi avemu doă-zecii și șapte de slove, și le împărțimu în Glasnice și Consunate.


Інтродъчере.

Гръматік'а есте ынъ мешешугъ прін каре ынвъцъмъ а ворві ші а скрі о лімбъ фъръ грешалъ.

Лімбъ нъмімъ ачелъ міжлокъ прін каре фачемъ кунощуте ші алтор'а ідеіле саъ кунощінцеле ноастре; астъ-фелъ сынт Ворбірea ші Скріерѣ.

Ка съ ворбімъ не служімъ ку ніще сунете саъ гласурі че ле пронунціъмъ ымпреунате, ші пе каре ле нъмімъ Зічері, кынд арътъ о фіінцъ саъ ынъ лукру, кум омъ, карте.

Ка съ скрімъ не служімъ ку Літеріле саъ Словеле, каре сынтъ ніще семне че арътъ сунетеле унеі лімбі, кум б, а, шчл.

Noі авемъ доъ-зечі ші шапте де слове, ші ле ымпърцімъ ын Гласніче ші Консунате.

Гласмічеле сынт шапте: а, е, і, о, у, ъ, ж, ші ле зічемъ астъ-фелъ, бъчі сінгуре ші фъръ ажуторулъ алтеі слове факъ ынъ сунетъ, ынъ гласъ.

Figura 3. Fragmentul textului original

Glasnicele sânt șapte: a, e, i, o, u, ă, â, și le zicemu astu'felu, căcii singure și fără ajutorulu altei slove facu unu sunetu, unu glasu.

CONCLUZII

Regulile descrise mai sus au servit drept bază la elaborarea algoritmului de dechirilizare a limbii române de tranziție.

Dechirilizarea prezintă un instrument care oferă mijloace de suport pentru specialiști în domeniul filologiei istorice și al conservării patrimoniului cultural.

Transformarea textului în grafia latină asigură aplicabilitatea metodelor de digitizare și a tehnicilor de cercetare a textelor istorice, bazate pe alfabetul latin.

În afară de suportul specialiștilor de cercetări istorice și filologice, instrumentul propus oferă acces liber la texte de importanță patrimonială, deoarece funcționează on-line și prezintă textele istorice într-o formă ușor de înțeles cititorului de astăzi.

Pe parcursul elaborării în continuare, instrumentul propus va fi completat cu mijloace de analiză ale textelor istorice.

Deoarece versiunea curentă a instrumentarului prezentat a fost aplicată cu succes la analiza diacronică manuală [7], realizarea modulului care efectuează această analiza în mod automat este pe primul plan. Implementarea acestui modul include atât aplicarea tehnicilor existente, cât și realizarea tehnologiilor proprii, care țin cont de specificul corpusului.

BIBLIOGRAFIE

1. Peneva, Juliana and Ivanov, Stanislav and Sotirova, Kalina and Doneva, Rossitza and Dobreva, Milena (2012) Access to Digital Cultural Heritage: Innovative Applications

of Automated Metadata Generation Chapter 1: Digitization of Cultural Heritage – Standards, Institutions, Initiatives. In: Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation. Plovdiv University Publishing House „Paisii Hilendarski", Plodvid, p. 25-67.

2. Cojocaru S., Boian E., Ciubotaru C., Colesnicov A., Demidova V., Malahov L. Regeneration of printed cultural heritage: challenges ang technolologies. Chișinău: The Third Conference of Matematical Society of the Republic of Moldova, 19-23 August, 2014, p. 481-489.

3. S. Cojocaru, L. Burtseva, C. Ciubotaru, A. Colesnicov, V. Demidova, L. Malahov, M. Petic, T. Bumbu, Ș. Ungur. On Technology for Digitization of Romanian Historical Heritage Printed in the Cyrillic Script, Proceedings of MFOI2016. Conference on Mathematical Foundations of Informatics, July 25-31, 2016, Chisinau, Republic of Moldova, p. 160-177.

4. Richard M. Reese. 2015. Natural Language Processing with Java. Packt Publishing.

5. Demidova Valentina. Particularităţile dechirilizării limbii române. In: Studia universitatis moldaviae, 2015, nr. 2(82), Seria „Ştiinţe exacte şi economice". Online 2345-1033, p. 16-20.

6. Demidova V. Particular Aspects of the Cyrillization Problem. Chișinău: The Third Conference of Matematical Society of the republic of Moldova , 19-23 August, 2014, p. 493-498.

7. D. Gifu, M. Dascalu, S. Trausan-Matu and L. K. Allen, „Time Evolution of Writing Styles in Romanian Language," 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), San Jose, CA, 2016, p. 1048-1054.

APORTUL INGINERIEI REPUBLICII MOLDOVA ÎN DEZVOLTAREA TEHNICII ȘI TEHNOLOGIILOR MODERNE (PARTEA II)

Academician **Ion BOSTAN**
Universitatea Tehnică a Moldovei

CONTRIBUTION OF REPUBLIC OF MOLDOVA ENGINEERING TO THE DEVELOPMENT OF TECHNIQUES AND MODERN TECHNOLOGIES (PART II)

Summary: This article describes, by specific examples, the contribution of Moldovan engineering to the development of techniques and modern technologies. An accomplishment, recognized on a global scale, is the invention of the Planetary Precessional Transmission in 1981 based on which, over three decades, various mechanisms and actuation systems for general and special purpose, comprising areas from cosmic flight technique to the submersible equipment, have been developed and implemented in manufacturing.

Through the material presented in this paper, the author aims to annihilate somehow the informational vacuum imposed during the Soviet period on the accomplishments of engineering sciences, which topics were interconnected with the military industry, or were considered as very important ones for the industrial and technological development of the country.

For the first time, the author shows a picture of the Moldovan engineering performance, exemplified by top achievements, obtained thanks to the efficiency of cooperation between national scientific researchers with the international structures empowered with research-innovation. The article highlights the scientific achievements of great value of some research teams of reference in engineering, which had a positive impact on the development and consolidation of engineering sciences and engineering, as a whole, at national and international levels.

Keywords: Engineering sciences, planetary precessional transmissions, satellite technologies, ferro-manganese concretions, renewable energy conversion systems, wind turbines.

Rezumat. Articolul reflectă, prin exemple concrete, aportul ingineriei Republicii Moldova în dezvoltarea tehnicii și tehnologiilor moderne. O realizare recunoscută la scară mondială este invenția, din anul 1981, a Transmisiei Planetare Precesionale în baza căreia, pe parcursul a trei decenii, au fost elaborate și implementate diverse mecanisme și sisteme de acționare cu destinație generală și specială cuprinzând domeniile de la tehnica cosmică de zbor până la cea submersibilă.

Autorul urmărește să suplinească vidul informațional impus în perioada sovietică asupra realizărilor științelor inginerești în materie de tehnică și tehnologii militare, precum și a cercetărilor considerate foarte importante pentru dezvoltarea industrial-tehnologică a țării.

În premieră, autorul prezintă un tablou al performanțelor ingineriei moldovenești, exemplificat prin realizări de vârf, obținute grație eficienței cooperării cercetătorilor științifici autohtoni cu structuri internaționale abilitate cu cercetarea-inovarea. În articol sunt scoase în evidență realizările de mare valoare ale unor echipe științifice de referință în inginerie care au avut un impact pozitiv asupra dezvoltării și consolidării științelor inginerești și a ingineriei în ansamblu în plan național și internațional.

Cuvinte-cheie: științe inginerești, transmisii planetare precesionale, tehnologii satelitare, concrețiuni fero-manganice, sisteme de conversie a energiilor regenerabile, turbine eoliene.

INTRODUCERE

Evoluția științelor inginerești în Republica Moldova în perioada 1970–1990 este marcată de extinderea tematicii investigațiilor științifice. Numai în cadrul Universității Tehnice a Moldovei (UTM), peste 50 de catedre s-au dezvoltat ca piloni de bază ai cercetării inginerești universitare și de pregătire a specialiștilor, inclusiv pe programe de doctorat și postdoctorat. Problematica cercetărilor efectuate în cadrul catedrelor corespundea, de regulă, profilului acestora, instituțiile de învățământ superior fiind stimulate astfel să sporească calitatea formării specialiștilor prin conexiunea cu cercetarea.

Rolul UTM în dezvoltarea științelor inginerești este important atât prin asigurarea diversității tematicii investigațiilor științifice, cât și prin cooperarea amplă cu instituțiile abilitate în domeniul cercetării, inclusiv ramurale (ISCR) și cu birourile specializate de proiectare constructiv-tehnologică (BSPCT).