

**GENOTYPDATA - špecializovaný informačný systém pre genetické zdroje  
rastlín**  
**GENOTYPDATA – specialized information system for plant genetic  
resources**

**Andrej URBANOVIČ<sup>1</sup>, Janka NÔŽKOVÁ<sup>1</sup>, Ján BRINDZA<sup>1</sup>**

<sup>1</sup>Inštitút ochrany biodiverzity a biologickej bezpečnosti, Slovenská poľnohospodárska univerzita v Nitre  
[akndrej.urbanovic@uniag.sk](mailto:akndrej.urbanovic@uniag.sk), [janka.nozkova@uniag.sk](mailto:janka.nozkova@uniag.sk), [jan.brindza@uniag.sk](mailto:jan.brindza@uniag.sk)

INFORUM 2005: 11. konferencie o profesionálnych informačných zdrojích  
Praha, 24. - 26.5. 2005

**Abstract:**

*GENOTYPDATA is a universal information system for record keeping, classification, cataloguing, study, selection, and evaluation of genetic resources individual plant species. Genotypes are evaluated by quantitative and qualitative traits. The system works according to intelligent system, which enables fast access to different numeral, textual and image data obtained from variability study. Possibility to apply any plant species according to unified methodology and same structure provides versatility and strategy of open information system.*

*For elaboration of GENOTYPDATA was used programming language Visual Basic 6.0. The basic structure of system consists from database system connected with mathematical and logical module. The system was set up for operation systems – Microsoft Windows 98, 2000 and XP. Information system is debugged for free settings, user's inputs and outputs. This provides maximal possibility of the system openness. For users the system provides fast and effectively to obtain process and use information for outputs in the area of plant genetic resources utilization.*

*In the paper is presented module – GENOTYPDATA Linum. In to the catalogue were included 121 genotypes of flax. They were evaluated in 39 quantitative and qualitative traits. Genotypes classification was realized according to innovated list of descriptors. The part of the information system is database of image documentation. It consists approximately from 1080 detailed images of leaf, flower, capsule, and seed.*

*Information system – GENOTYPDATA has practical utilization in the area of plant breeding, and for recording, evaluation and cataloguing of genetic resources in gene banks. Its possibility for use is also in an agriculture production, research and educational process.*

## Úvod

Informácie dnes zaraďujeme vedľa materiálových, energetických a finančných zdrojov k hlavným faktorom podmieňujúcim pokrok nielen v technike, ale vo všetkých oblastiach ľudskej činnosti. Nie náhodou v dnešnej dobe hovoríme o probléme informačnej explózie. Úspešne čeliť tomuto novému fenoménu je možné použitím efektívnych spôsobov získavania údajov o stave reálneho sveta a ich racionálne ukladanie, vyhľadávanie alebo využitie odvodených dát a informácií, čo skvalitňuje rozhodovacie procesy (Urbanovič 2005).

Prvkami informačného systému sú informácie. Tieto prvky informačného systému majú tvar závislý na užívateľovi informačného systému a na funkciu informačného systému.

Informačnými technológiami nazývame postupy, algoritmy a metódy, ktorými je možné v nadviazaní na technickú infraštruktúru informatiky efektívne a kvalitne vykonávať operácie s veľkým množstvom dát a podporiť tak proces zvyšovania pragmatickej hodnoty odvodzovaných dát a informácií, ktoré je možné z dát vyťažiť (Urbanovič 2005, Moos 1993).

Za účelom efektívneho využívania informácií sa v podnikoch, organizáciách a inštitúciách používajú systémy, ktorých úlohou je produkovať kvalitné a relevantné informácie potrebné pre riadiaci proces. Základnou funkciou informačných systémov je teda získať, spracovať a poskytnúť informácie užívateľom v čase, keď ich potrebujú a to v požadovanej kvalite (Moos 1993, Pechr 2003).

Rozvoj informačnej a komunikačnej techniky vo svete otvára možnosti pre vytvorenie komplexných systémov ako sú katalógy, špecializované databázy, informačné systémy pre evidenciu a charakterizáciu genetických zdrojov rastlín pre výživu a poľnohospodárstvo. Okrem evidovania niekoľko desiat tisíc genetických zdrojov v rôznej forme, poskytujú kapacitu aj na uchovávanie tisícov údajov získaných z evidencie, hodnotenia a charakterizácie jednotlivých zdrojov (Brindza et al. 1998).

V pôsobnosti každej krajiny je snaha o vytvorenie národných ale aj medzinárodných databáz pre genetické zdroje (Pavelek 1994, Puzone 1998, Hintum 1997). Avšak s príliš veľkým a stále sa zväčšujúcim počtom vzoriek vo svetových kolekciách genetických zdrojov narastá problém ich kompletného hodnotenia a poznania ich špecifických znakov (Diederichsen, Richards 2001). Nedostatočné množstvo, nízka hodnovernosť údajov získaných zo štúdia vnútrodruhovej variability a tak isto aj nejednotnosť v systéme evidencie a hodnotenia neprináša skutočný pohľad na stav kolekcií genetických zdrojov vo svete (Diederichsen, Raney 2001). Informácie, ktoré centrálné databázy pre kultúrne druhy rastlín (národné, medzinárodné) poskytujú sú obmedzené na pasportné údaje, údaje o prostredí, manažmente (IPGRI 2004), pretože sú najviac uniformné pre všetky plodiny. Údaje z hodnotenia a charakterizácie (špecifické vlastnosti a znaky) genetických zdrojov sú viac požadované a zaujímavé pre širší okruh užívateľov. Avšak variabilnosť v týchto údajoch je veľmi veľká a sú rozdielne pre každú plodinu. Preto vzniká potreba vytvorenia univerzálneho a unifikovaného systému, ktorý umožňuje vkladanie aj takýchto údajov a informácií podľa unifikovanej metodiky a rovnakej štruktúry. Charakter vkladanych údajov môže byť rôzny podľa potrieb užívateľa.

V roku 1992 sa na katedre genetiky a šľachtenia rastlín Slovenskej poľnohospodárskej univerzity v Nitre začalo s tvorbou špecializovaného informačného systému pre genetické zdroje kultúrnych druhov rastlín, *GENOTYPDATA*, ktorý poskytuje, triedi a spracováva informácie podľa potrieb užívateľa (Brindza et al. 1996, Brindza et al. 1998). Prvé dva moduly *GENOTYPDATA Fragaria* a *GENOTYPDATA Pisum* boli uzavreté. Posledný tretí modul (*GENOTYPDATA Linum*), ktorý je aplikovaný na kolekciu ľanu siateho je otvorený pre nové vstupy zo strany užívateľov.

## **Materiál a metódy**

Podkladový materiál pre vývoj tretieho modulu informačného systému (*GENOTYPDATA Linum*) bol tvorený súborom genotypov (n=121) ľanu siateho (*Linum usitatissimum* L.). Získané boli z Národnej génovej banky Slovenska (56 genotypov) a z výskumného ústavu Agritec s.r.o. Šumperk (65 genotypov). Pasporné údaje o jednotlivých genotypoch boli sústredené z uvedených inštitúcií. Genotypy sa zhodnotili vybranými kvantitatívnymi a kvalitatívnymi znakoch vo vegetačných rokoch 2002 – 2004. Údaje z hodnotenia sa analyzovali matematicko-štatistickými metódami. Podklady pozostávali aj z obrazovej dokumentácie detailných záberov listu, kvetu, korunných lupienkov, kalicha, tobolky a semena ľanu siateho (Nôžková 2005).

Pre charakterizáciu a klasifikáciu genotypov sa použil inovovaný klasifikátor (Nôžková 2005). Nejednotnosť v systéme hodnotenia existujúcich medzinárodných klasifikátorov (UPOV 1995; Pavelek, Faberová 2000; Rosenberg 1978; Kutuzova et al. 1989; IFDB 1994) bola hlavným dôvodom inovácie klasifikátora. Tým bol vytvorený unifikovaný systém hodnotenia a charakterizácie pre všetky genotypy ľanu siateho.

Informačný systém *GENOTYPDATA* je naprogramovaný ako databázový informačný systém tretej generácie „Relačné databázy“ a štvrtej generácie „Objektovo – orientovanej databázy“.

Bol naprogramovaný v programovacom jazyku Visual Basic 6 a ako databázová jednotka bola použitá databáza softwaru Microsoft Access. Program spolupracuje s jedným externým softwarom „Acrobat Reader“.

Informačný systém *GENOTYPDATA* je odladený pre operačný systém Windows 98, 2000 a XP.

## **Výsledky**

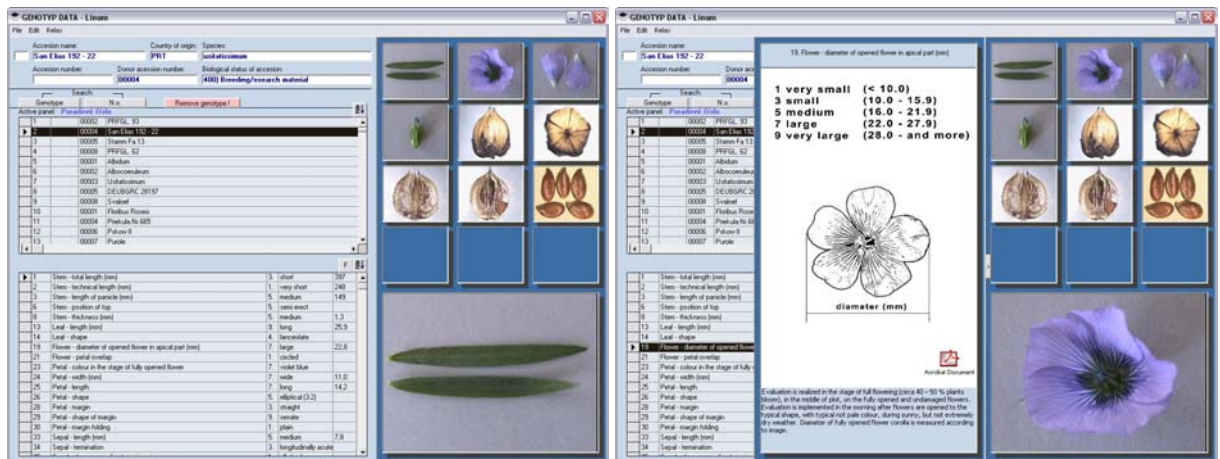
Základnými prvkami informačného systému (IS) *GENOTYPDATA Linum* sú informácie o jednotlivých genotypoch ľanu siateho. Prvky informačného systému majú tvar závislý na užívateľovi informačného systému a na funkciu IS. V našom prípade sa rozdeľujú na dve základné skupiny informácií (schéma). Prvú skupinu tvoria pasporné údaje, pomocou ktorých sa zabezpečuje základná evidencia genotypov vo všetkých svetových inštitúciách. Pre zabezpečenie univerzálnosti IS sa zvolilo 33 pasporných deskriptorov EURISCO (2002), ktoré oficiálne platia pre všetkých, ktorí uchovávajú a evidujú genetické zdroje

Druhú skupinu informácií o genetických zdrojoch tvoria údaje z hodnotenia a charakterizácie (schéma). Rozmanitosť a špecifickosť týchto údajov je veľmi veľká a je náročné ich zosúladiť a zjednotiť do jedného funkčného systému. Hľadali sa vzájomné vzťahy medzi jednotlivými prvkami, ktoré do systému vstupovali. Tieto vzťahy plnili rôzne funkcie.

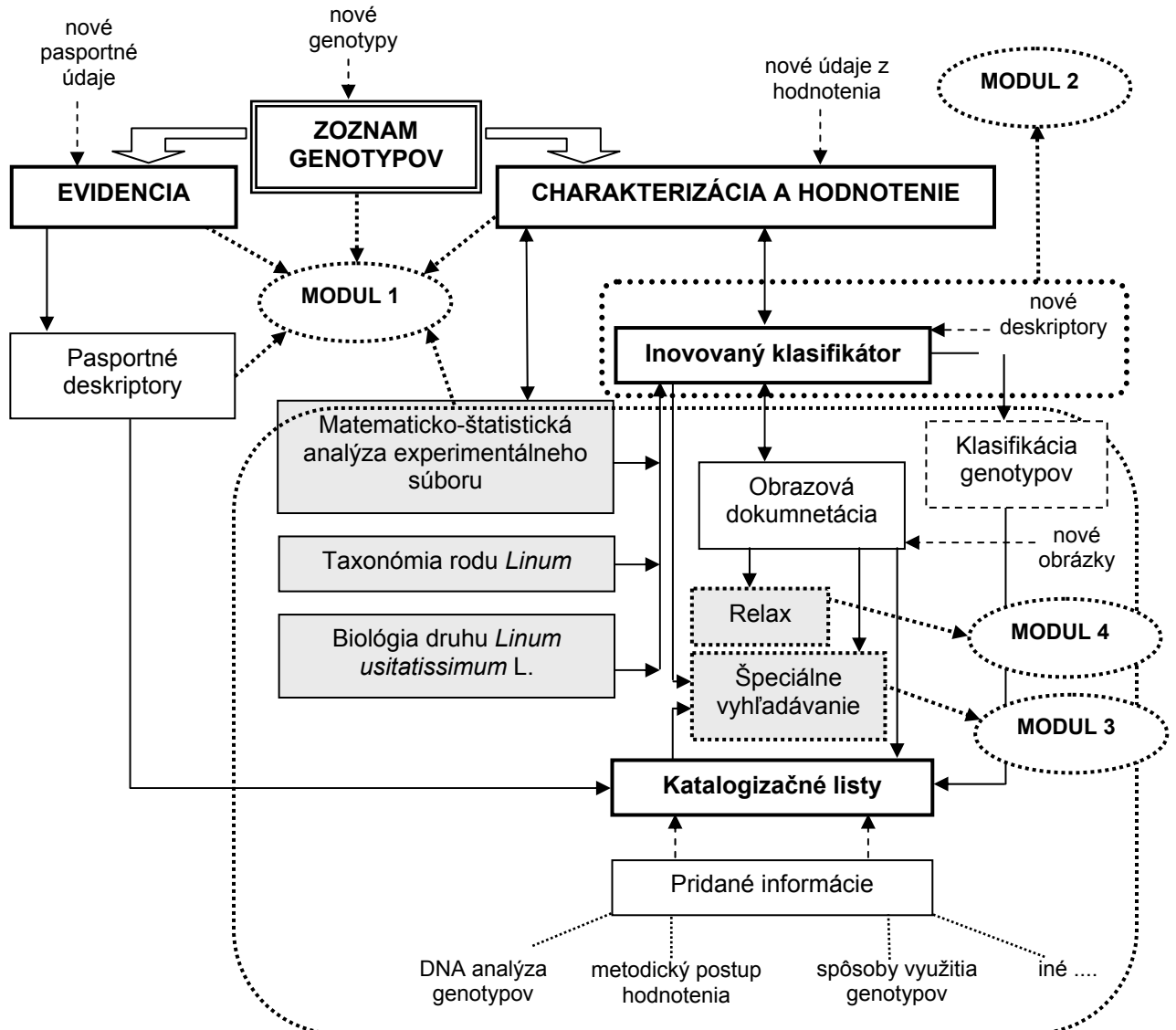
Zosúladenie všetkých informácií o genotypoch ľanu siateho sa zabezpečilo vytvorením funkčnej štruktúry IS. Založená je na základnom databázovom module, ktorý tvorí jadro informačného systému. Toto jadro predstavuje katalóg genotypov vybraného rastlinného druhu (modul 1 – schéma). Sú tu uložené všetky údaje o jednotlivých genotypoch a informáciách prislúchajúcich ku genotypom. Na katalógovom liste vybraného genotypu (výstup z databázy katalóg) sa nachádzajú pasporné údaje, údaje z hodnotenia a charakterizácie, klasifikácie, obrazová dokumentácia (obr. 1).

Tento základný modul poskytuje užívateľom základné prvky databázových informačných systémov, a to ukladanie údajov, triedenie, vyhľadávanie a zoradovanie údajov v informačnom systéme. Sú v ňom možnosti práce s textovými dokumentmi, pdf dokumentmi a multimediálnymi dokumentmi (obr. 2 – prílohy).

**Obrázok 1:** Katalógový list vybraného genotypu s účelným prepojením na deskriptor znaku



**Schéma:** Základná štruktúra informácií a modulov informačného systému - GENOTYPDATA



Hlavnou myšlienkou však bolo navrhnuť tento systém ako otvorený a prístupný pre široký okruh užívateľov. Otvorenosť systému by však nemala narušiť jeho funkčnosť aj napriek veľkej variabilite kultúrnych druhov rastlín. Preto bol náš systém upravený tak aby bola možnosť zmeny jednotlivých modulov. To znamená, že systém nie je otvorený len na pridávanie nových genotypov do zoznamu (obr. 3 – prílohy) a nových údajov o nich (schéma).

Druhou časťou celkovej štruktúry IS bol ďalší nosný databázový modul pomocou ktorého sa realizujú všetky procesy týkajúce sa hodnotenia a charakterizácie genotypov. Je to inovovaný klasifikátor (modul 2 – schéma), ktorý bolo nutné postaviť tak aby bol plne voľne upraviteľný užívateľom. Umožňuje mu nadefinovať základný prvok (deskriptor) potrebný pre klasifikáciu a charakterizáciu genotypov (obr. 4).

**Obrázok 4:** Formulár pre nadefinovanie nového deskriptora do celkového zoznamu modulu 2 - klasifikátor

Univerzálnosť a jednotnosť klasifikátorov pre rôzne rastlinné druhy spočíva v zhodnej štruktúre jednotlivých deskriptorov, ktoré majú tri základné časti – bodové kódovanie úrovní, slovná charakteristika úrovní a v prípade kvantitatívnych znakov intervaly. Maximálny počet úrovní deskriptora je 9. Minimálny počet úrovní, pri alternatívnych znakoch (1 nie, 9 áno), je 2. Táto vlastnosť klasifikátora umožňuje užívateľovi IS upravovať a vkladať nové informácie o genotypoch, podľa svojich potrieb. Tento modul po nadefinovaní predstavuje základný pracovný prvok informačného systému GENOTYPDATA. Vyriešením otvorenosti a spracovania nadefinovaného modulu 2 - Klasifikátor sme sa pokúsili vyriešiť dlhoročné problémy s flexibilitou klasifikačných systémov pre široké spektrum užívateľov pracujúcich vo sfére genetických zdrojov.

K týmto dvom základným databázovým modulom sme v informačnom systéme pridali ešte dva moduly, ktoré posúvajú tento informačný systém ďalej v spracovaní vstupných údajov do systému. Je to modul 3 – špeciálne vyhľadávanie podľa zadaných kritérií a modul 4 s názvom „Relax“ (schéma).

V module pre špeciálne vyhľadávanie je umožnené užívateľovi pomocou zadaných kritérií (modul 2 – klasifikátor) spracovať uložené údaje (modul 1 – katalóg) a následne vyhľadať a poskytnúť užívateľovi výstupné informácie (obrazová dokumentácia) potrebné v rozhodovacom procese klasifikácie genotypov a šľachtenia nových odrôd (obr. 5 - prílohy).

Jednou z najväčších novínok v informačných systémoch pre genetické zdroje je naprogramovanie a vloženie modulu 4 - Relax do štruktúry IS GENOTYPDATA. Tento modul predstavuje myšlienku vizuálneho spracovania variability biologického materiálu (obr. 6 - prílohy). Modul „Relax“ je zostavený na rozhodovacích a matematických operáciách náhodného zobrazovania a vizuálneho spracovávanía multimediálnych údajov. Poskytuje užívateľovi iný pohľad na široké spektrum farieb, tvarov jednotlivých častí rastliny a tým aj pohodlnejšie štúdium a odhaľovanie nových, nepoznaných krás biologickej diverzity. Pohľad na veľké množstvo jednotlivých prvkov obrazovej dokumentácie napomáha užívateľovi aj ľahšie rozhodovanie v hodnotení genetických zdrojov.

## Diskusia

Výber deskriptorov, ktoré by mali byť zahrnuté do databázy vybranej plodiny závisia od typu riešenej otázky ale aj od typu opisovanej plodiny. Klasifikátory sú potom často spracované podľa výskumného tímu, ktorý z danou plodinou pracuje. Takto vznikajú medzi databázami rozdiely týkajúce sa formy a výberu deskriptorov (Hulden 1997).

Špecificky spracované klasifikátory slúžia iba úzkemu okruhu užívateľov a pre záujemcu z inej oblasti sú nepoužiteľné. Pre zabezpečenie štandardizácie centrálnych databáz pre genetické zdroje rastlín sa preto vo väčšine prípadov využívali pasportné deskriptory, deskriptory o prostredí a manažmente. Pasportné údaje o genotypoch ľanu siateho sú úzko orientované iba na úzku skupinu užívateľov z radov – pracovníkov v génových bankách (kurátori), šľachtiteľov a výskumných pracovníkov, ktorí hľadajú konkrétny materiál pre realizáciu svojich úloh (Horváth, Szabó 1997). Deskriptory pre hodnotenie a charakterizáciu sa pre svoju veľkú variabilnosť a špecifickosť k jednotlivým plodinám mnoho krát do databáz nedostali (Horváth, Szabó 1997).

Informačný systém GENOTYPDATA sa z časti pokúsil riešiť otázku nejednotnosti pri hodnotení a charakterizácii genetických zdrojov. Systém je otvorený a upravený pre nové vstupy zo strany užívateľa, dokonca umožňuje zmeny jednotlivých modulov. Takto upravený informačný systém umožňuje využívať genetické zdroje ako existenčné, východiskové, produkčné, študijné, informačné a inšpiračné zdroje.

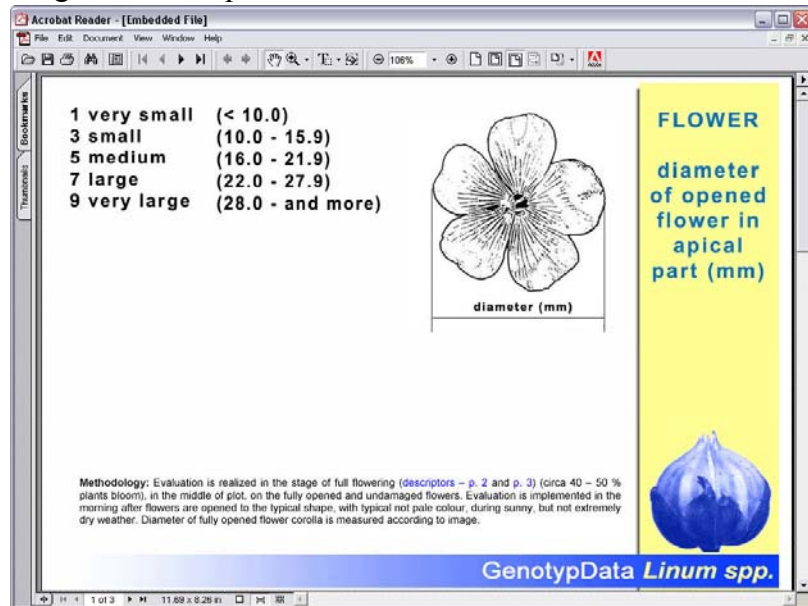
Databázy s obrazovou dokumentáciou predstavujú nové možnosti pri práci s genetickými zdrojmi, pretože vizuálne údaje sú vhodnejšie pre priame hodnotenie a pre ďalšiu elektronickú analýzu (Fukui a Kakeda, 1990, Blaich, Blaich, 1990, Brindza, 1992, 1996, Mincione et al. 1994). Zaradením dvoch nových modulov do štruktúry systému – relax a špeciálne vyhľadávanie (modul 3 a 4), ktoré využívajú obrazovú dokumentáciu, sa sprístupnil užívateľovi pohodlnejší, expresnejší a exaktnejší prístup k informáciám v procese poznávania, šľachtenia, hľadania nových možností využitia a odkrývania dosiaľ nepoznaných krás biologickej variability.

### **Použitá literatúra:**

- BRINDZA, J. – BALOGH, Z. – GAŽO, J. – ZIMMEROVÁ, J. – STEHLÍKOVÁ, B. (1996): Vývoj novej verzie informačného systému GENOTYPDATA. In: Genetické zdroje rastlín – ročenka. SPU Nitra, 1996, s. 116 – 121. ISBN 80-7137-422-9
- BRINDZA, J. – BALOGH, Z. – ZIMMEROVÁ, J. - GAŽO, J. (1998): GENOTYPDATA – computerized information system for documentation and evaluation of genetic resources. In: Characterization and documentation of genetic resources utilizing multimedia databases. Neapol – workshop : IPGRI, 1998, s. 22 – 27. ISBN 92-9043-388-4
- DIEDERICHSEN, A. – RANEY, J. P. (2001): Flax Germplasm – Characterization and Screening for Seed-Oil Quality – final report for the project. Kanada- Saskatoo Research Centre, s. 23.
- DIEDERICHSEN, A. – RICHARDS, K. W. (2001). Cultivated flax and genus *Linum* L. – Taxonomy and germplasm conservation. In: MUIR, A. D. – WESTCOTT, N. D.: Flax – the genus *Linum*. Hrawood Acad. publ., Amsterdam, 2001, s. 22 – 54.
- EURISCO descriptors – for uploading information from National Inventories to EURISCO (2002): [http://www.ecpgr.cgiar.org/epgris/Tech\\_papers/EURISCO\\_Descriptors.doc](http://www.ecpgr.cgiar.org/epgris/Tech_papers/EURISCO_Descriptors.doc) (7.3.2005)
- HINTUM van, J.L. (1997): Central crop databases – an overview. In: Central Crop Databases: Tools for Plant Genetic Resources Management. Report of workshop – ECP/GR Networks, Hungary, 1996, p. 17 – 19. ISBN 92-9043-320-5
- IPGRI and COMAV. (2004): Descriptors for Pepino (*Solanum muricatum*). International Plant Genetic Resources Institute, Rome, Italy, and Centro de Conservación y Mejora de la Agrodiversidad Valenciana, Valencia, Spain. ISBN 92-9043-616-6
- MOOS, P. 1993. Informační technologie. Praha: ČVUT, 1993, 200 s. ISBN 80-01-01048-1
- NÔŽKOVÁ, J. (2005): Hospodárska významnosť a možnosť ďalšieho využitia udržiavaného genofondu ľanu siateho (*Linum usitatissimum* L.) – dizertačná práca. SPU Nitra, 140 s.
- PAVELEK, M. (1993-1994): Vytváření evropské databáze lnu. In: Genetické zdroje rastlín – ročenka. SPU Nitra, 1993-1994, s. 117-118. ISBN 80-7137-218-8
- PECHR, B. 2003. Využitie informačných technológií v manažmente výrobných poľnohospodárskych technológií. In: Súčasný stav a výhľad technického a technologického zabezpečenia pracovných procesov v poľnohospodárstve SR. Agentúra slovenskej akadémie pôdohospodárskych vied Nitra, 2003, s. 53-55, ISBN 80-968665-6-7
- PUZONE, L. (1998). A software prototype for germplasm image databases. In: Characterization and documentation of genetic resources utilizing multimedia databases. Neapol – workshop : IPGRI, 1998, s. 28 – 35. ISBN 92-9043-388-4
- URBANOVICĎ, A. 2005. Využitie informačného systému v manažmente techniky pri pestovaní repky olejnej - dizertačná práca. SPU Nitra, 2005, 185 s..

## Prílohy:

**Obrázok 2:** Príklad metodického postupu hodnotenia ako pridaná informácia v module 1 – katalóg vo formáte pdf.



**Obrázok 3:** Formulár pre nadefinovanie nového genotypu do celkového zoznamu v module 1 – katalóg

**Add genotype:**

**Enter information:**

Accession name:  Acquisition date:  Latitude of collecting site:  Donor institute code:   
Accession number:  Country of origin:  Longitude of collecting site:  Location of safety duplicates:   
Donor accession number:  Collecting date of sample:  Elevation of collecting site:  Breeding institute code:   
Species authority:  Species:  Institute code:   
Subclass:  Genus:  Collecting number:   
Subclass authority:  Ancestral data:  Collecting institute code:   
Common crop name:  Other identification (numbers) associated with the accession:   
Location of collecting site:  Type of germplasm storage:  Remarks:   
Collecting/acquisition source:   
Biological status of accession:

**Enter information:**

Leaf:  Capsule cut 1:   
Flower:  Capsule cut 2:   
Petals:  Seed:   
Calyx:  Free no.10:   
Capsule - horizontal:  Free no.11:   
Capsule - vertical:  Free no.12:

**Enter information:**

ID	Descriptor's name	Value
1	Stem - total length (mm)	
2	Stem - technical length (mm)	
3	Stem - length of panicle (mm)	
6	Stem - position of top	
8	Stem - thickness (mm)	
12	Leaf - length (mm)	

Descriptor's values:

Number of descriptor:   
Status descriptor:   
Value of descriptor:   
Remark of descriptor:

Attach document to descriptor:    
Attach image description to descriptor:



Obrázok 5: Modul - špeciálneho vyhľadávania



Obrázok 6: Modul - relax

