

# Národní geovědní bibliografie GeoLib

National Earth sciences bibliography

HANA BREITEROVÁ

Česká geologická služba, Klárov 3, 118 21 Praha 1;  
hana.breiterova@geology.cz

*Key words: geology, bibliography, databases, applied informatics, cooperation*

**Abstract:** The new information infrastructure in the field of the Earth sciences is based on the utilization of data for the Registry of Information on the Results of scientific and research work (abbreviation RIV in Czech) and other publication activities of geo-scientific organizations. The base enable to searching records through indexes and doing specific retrieval.

Výsledkem řešení projektu VaV SP II 4h3/22/07 Národní geovědní bibliografie (dále NGB) je vytvoření nové informační infrastruktury v oblasti věd o Zemi. Tato infrastruktura vznikla využitím dat pro Registr informací o výsledcích (dále RIV) a další publikační činnosti pracovníků významných organizací zabývajících se vědami o Zemi a jejich spojením s Geologicko-mineralogickou bibliografií ČR, zpracovávanou v knihovně České geologické služby. Propojení dosud samostatných databází umožní uživateli získat informace v jednotném informačním rozhraní. Příspěvateli do Národní geovědní bibliografie jsou Přírodovědecká fakulta Univerzity Karlovy v Praze, Přírodovědecká fakulta Masarykovy univerzity, Národní muzeum, Moravské zemské muzeum, Geologický ústav AV ČR, v. v. i., Geofyzikální ústav AV ČR, v. v. i., Ústav struktury a mechaniky hornin AV ČR, v. v. i., a Česká geologická služba (ČGS), která celý projekt VaV řídila.

## Harmonizace a analýza

V průběhu celého období řešení projektu byla věnována pozornost harmonizaci dat z nesourodých databází, a to jak v datovém skladu ČGS, tak u externích organizací. Nezbytným předpokladem pro zdárné vyřešení projektu bylo provedení datové analýzy. Důležité bylo především stanovení používaných formátů dat, retrospektivy a v neposlední řadě též zjištění toho, kdo záznamy eviduje, popř. zda probíhá nějaká následná kontrola. Tyto skutečnosti byly důležité zejména proto, že pravidla pro zápis údajů v jednotlivých formátech se liší, s čímž bylo třeba počítat v návrhu vytvářené aplikace.

Harmonizace dat ČGS byla zajišťována jejich úpravou pomocí uživatelských rozhraní (aplikací) a dále přímo na fyzické úrovni databází. Proces harmonizace dat probíhal průběžně jak v bibliografické článkové databázi knihovny, tak v datovém skladu GeoPub, které jsou zdrojem dat pro NGB. Harmonizace dat v databázích, které nejsou zdrojem pro NGB (separáty, archivní materiály) byla nutná, aby data v NGB a data v ostatních bibliografických databázích ČGS byla kompatibilní a připravená pro případně budoucí jednotné vyhledávání v obou systémech současně.

Byly zjištěny tři různé formáty: formát UNIMARC –

Akademie věd, bibliografie České geologické služby, formát MARC 21 – databáze Přírodovědecké fakulty MU, formát podle metodiky RIV – Národní muzeum, Moravské zemské muzeum, Přírodovědecká fakulta UK a databáze GeoPub. Formáty UNIMARC a MARC 21 jsou běžně používané knihovní standardy, konverze mezi nimi je dořešená, ale bylo třeba vyřešit konverzi formátu RIV.

Retrospektiva záznamů v databázích je různá, největší má bibliografie ČGS – zpět do roku 1986. Záznamy z Akademie věd jsou doplněny od roku 1993, někde od 1994. Přírodovědecká fakulta UK uvádí retrospektivu od r. 1998, Národní muzeum několik let, ostatní organizace se k retrospektivě nevyjádřily, nicméně na webových stránkách Přírodovědecké fakulty MU je možno vyhledávat publikace od roku 1993. Databáze GeoPub obsahuje záznamy od roku 1999.

Šíře evidovaných publikačních aktivit je velmi rozdílná. Z analýzy vyplynulo, že všechny organizace bez rozdílu evidují výstupy pro RIV. Moravské zemské muzeum a Přírodovědecká fakulta MU evidují i další výstupy, ne však všechny. V těchto organizacích je na uvážení autorů, které další výstupy zadají do evidence. Naopak v ústavech Akademie věd, Přírodovědecké fakultě UK a v ČGS jsou evidovány veškeré výstupy a aktivity vědeckých a výzkumných pracovníků.

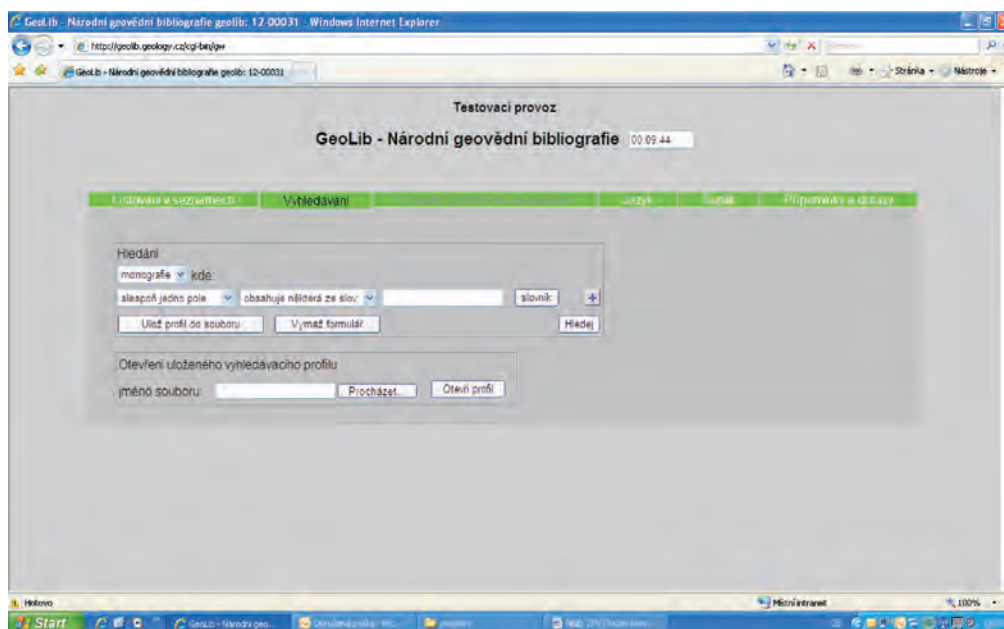
## Metody systémů získání a zobrazení dat

Pro získávání a zobrazení dat existují tři možná řešení:

1. centrální – data jsou ukládána do centrálního úložiště a dávkově aktualizována. Vyhledání a konsolidace dat by probíhaly v ČGS v rámci stávajícího technologického vybavení (datový a aplikační server Oracle);
2. distribuovaná – data zůstanou v původních databázích (u poskytovatelů) a systém – centrální aplikace v ČGS – do nich pouze nahlíží, což vyžaduje těsnější propojení (techniky i kapacitně) se spolupracujícími organizacemi. Toto řešení vůbec neřeší duplicitní záznamy, ale na druhou stranu umožní spolupracujícím organizacím přímo sdílet zdrojová data bez dalších kapacitních nároků;
3. kombinace obou popsáných metod, založená na technologických a personálních možnostech jednotlivých poskytovatelů.

Tabulka 1. Statistika záznamů podle příspěvatele

příspěvatele	formát	počty záznamů (stav k 11. 11. 2010)			
		celkem	převedeno	monografie	články
Česká geologická služba	UNIMARC	24 536	24 535	0	24 489
	RIV	4 449	4 349	385	3 941
AV ČR – Geologický ústav	UNIMARC	4 156	3 821	216	3 592
AV ČR – Geofyzikální ústav	UNIMARC	3 117	2 793	226	2 516
AV ČR – ÚSMH	UNIMARC	2 968	2 375	144	2 220
UK Praha – Přírodovědecká fakulta	RIV	1 209	1 209	13	1 188
MU Brno – Přírodovědecká fakulta	RIV	904	903	61	837
Národní muzeum, Praha	RIV	61	61	1	58
Moravské zemské muzeum, Brno	RIV	23	23	0	23
součet		41 423	40 069	1 046	38 864
celkem po automatické deduplikaci				1 046	38 695



Obr. 1. Webové rozhraní testovacího provozu aplikace GeoLib.

Každá z metod má své výhody a nevýhody a každá předpokládá dlouhodobé personální zajištění. Z diskuse s partnery, kteří nyní dodávají data do NGB, i z vlastních zkušeností vyplynulo, že kapacitně schůdnější bude zajistit pravidelné posílání dat do centrálního skladu, než zajistit kapacitu dbající na kompatibilitu všech úrovní v případě distribuované metody. Podle zadání projektu hrálo zcela klíčovou úlohu řešení duplicit nebo multiplicit záznamů, které vznikají kooperací autorů napříč poskytovateli. S ohledem na tato fakta byla po důkladném zhodnocení vybrána metoda centrální.

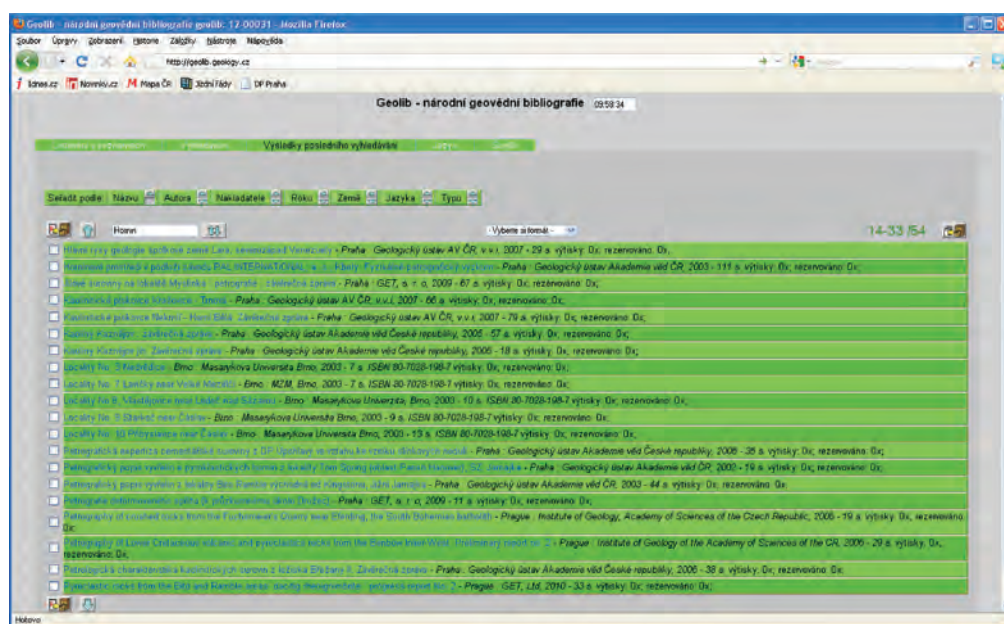
## Popis aplikace GeoLib

Aplikace pro sběr dat z oblasti geologie GeoLib je vyvinuta v prostředí databázového stroje TINMAN. Jde o aplikaci typu klient-server, která zpracovává jak interní data vytvořená ČGS, tak i data dalších spolupracujících institucí. Aplikace

GeoLib předpokládá data ve dvou různých formátech (UNIMARC, RIV) a je navržena tak, aby umožňovala snadné přidání dalších spolupracujících knihoven.

Sběr dat od jednotlivých příspěvatele probíhá manuálně, vlastní naplňování databáze je však zcela automatizované. Aplikace GeoLib periodicky kontroluje, zda na server nebyly umístěny nové soubory, které se mají zpracovat. V takovém případě pak spustí import záznamů, jejich indexaci a automatickou deduplikaci (sloučení duplicitních záznamů). Ta probíhá na základě deduplikačního klíče, odvozeného z nastavení definovaného správce databáze. Součástí klíče může být standardní číslo (ISBN, ISSN) v kombinaci s dalšími údaji – typem záznamu (monografie, článek), částí názvu dokumentu + částí jména autora + rokem vydání (+ u článků navíc ještě ročníkem a číslem). Zároveň vygeneruje statistiky importu a celkové statistiky, které volitelně odešle e-mailem na předdefinované adresy. Statistika záznamů v datovém skladu k 11.11.2010 je zobrazena v tab. 1.

Obr. 2. Výsledky vyhledávání podle jména autora.



Pro každého přispěvatele se tato aplikace spouští s jiným přihlášením (tzv. autologinem), což umožňuje zachovat informace o přispěvateli a formátu záznamů a zároveň v závislosti na nastavení importních tabulek směřovat převod údajů věcného popisu. Pro všechny přispěvatele se spouští stejný importní skript, avšak s jiným nastavením.

Administrátor z ČGS má přístup do aplikace GeoLib prostřednictvím klientské části, která umožňuje provádět manuální deduplikaci a opravy záznamů. Tyto opravy je však doporučeno provádět pouze výjimečně, protože mohou být přepsané při případné aktualizaci záznamu. Doporučený postup v tomto případě je provést opravu přímo ve zdrojové databázi a následně odeslat záznam k aktualizaci do aplikace GeoLib.

Koncoví uživatelé mají přístup do aplikace GeoLib prostřednictvím www rozhraní GeoWeb (obr. 1). Toto rozhraní umožňuje listování v rejstřících, vyhledávání podle různých kritérií, řazení výsledků vyhledávání, uložení vyhledaných záznamů i vyhledávacích dotazů. Při zadávání dotazu lze použít Booleovské operátory (AND, OR) a k dispozici jsou rovněž autoritní seznamy, popř. číselníky pro jednotlivá pole. Vyhledané záznamy lze zobrazit ve dvou různých formátech a exportovat je do xml. Výsledky vyhledávání podle jména autora jsou na obr. 2.

## Závěr

Propojením různých systémů evidence publikační činnosti vědeckých pracovníků mezi organizacemi čistě výzkumnými, vzdělávacími a servisními, které se zabývají vědami o Zemi, vznikl ucelený systém s uživatelsky příjemným vyhledávacím prostředím. Systém současně vytvořil dobré prostředí pro dlouhodobou vzájemnou spolupráci mezi zúčastněnými organizacemi.

Nově vytvořená Národní geovědní bibliografie se posléze stane součástí informačního systému Ministerstva životního prostředí (portálu životního prostředí) jako její specializovaná část (tematická doména). Národní geovědní bibliografie je volně přístupná na webových stránkách. Po dokončení testovacího provozu a sladění s portálem ČGS (červen 2011) bude k dispozici i na stránkách kooperujících organizací, MŽP a Národní knihovny, se kterou knihovna České geologické služby dlouhodobě spolupracuje.

Národní geovědní bibliografie přináší uživatelům možnost jednotného vyhledávání informací o výsledcích výzkumu v oblasti věd o Zemi, které vznikly v různých výzkumných institucích. V budoucnu by se mohla stát dobrým základem pro vytvoření Národní geovědní digitální knihovny.

*Poděkování. Práce byly provedeny v rámci projektu VaV MŽP SP II 4h3/22/07.*