

Digitálny archív na www.geology.sk

RNDr. Štefan Káčer

Štátny geologický ústav Dionýza Štúra, Mlynská dolina 1
817 04 Bratislava, Slovenská republika
stefan.kacer@geology.sk

Abstrakt.

Archív Geofondu je jedinečnou archívnu zbierkou geologických správ a posudkov vzniknutých z geologického bádania na území Slovenskej republiky. Jeho povinnosťou je zhromažďovať, ochraňovať a sprístupňovať vyše 87 000 záverečných správ, čo je asi 1,5 kilometra textového a grafického materiálu. Najstaršia správa je z roku 1874. Archív je verejne prístupný a nachádza sa v budove ŠGÚDŠ v Bratislave.

V rámci riešenia projektu geologickej úlohy **Geologický informačný systém GeoIS** sme prepracovali a sprístupnili existujúcu bibliografickú databázu archívu Geofondu. Implementáciou produktu Webom (internetová nadstavba IBM DB2 Content Manager – správa neštruktúrovaného obsahu dokumentov) bola poverená na základe výsledkov verejného obstarávania firma YMS, a.s. Trnava. Zdefinovali sme požiadavky na funkčnosť systému, stanovenie vlastností atribútov, položiek, väzby medzi nimi, hodnoty číselníkov, výstupné tlačové formáty, možnosti vkladania digitálne spracovaných dokumentov a fulltextové vyhľadávanie s reálnou možnosťou sprístupnenia informácií externým užívateľom. V rámci migrácie predchádzajúcej verzie bibliografických údajov sme vykonali generálne opravy chybných dát v databáze, zjednocované údaje aj z predchádzajúcich systémov, doplnenie zničených údajov z nekonzistentných alebo inak poškodených záznamov.

V digitálnom archíve Geofondu sú registrovanému bádateľovi dostupné naskenované textové a niektoré prílohy časti záverečných správ v PDF formáte. Extrahovaním textov záverečných správ prostredníctvom aplikácie na optické rozpoznávanie znakov (OCR) s následným uložením do centrálnej databázy systém umožňuje vyhľadávanie kľúčových slov priamo v textoch záverečných správ.

1.4.2008 bol Digitálny archív Geofondu sprístupnený verejnosti. Vstup a registráciu bádateľ nájde na oficiálnej stránke Štátneho geologického ústavu Dionýza Štúra - www.geology.sk.

Kľúčové slová: mapový server, dátový model, geodatabáza.

Abstract.

Archive of Geofond

The SGUDS activity is focused in:

- solution of projects of geological research and exploration
- creation and use of information system in geology
- **registration, collecting, evidence and making available of results of geological works** carried out within the Slovak Republic territory

The Geofond is a modern archive of geological documentation in the sphere of ME SR
The Archive of SGUDS is a unique fund of the geological works results covering the entire territory of the Slovak Republic

An open study room supplies the investigators with archive documents and other archive facilities for study purposes

- In the past the report search was realized through catalogue index cards
- On the internet website signal data on final reports were placed
- This list was actualized once a month
- The Archive employees prepare upon request various recherche from stored reports
- The Geofond Archive contains reports, which have been delivered since 1887
- The Archive keeps more than 87 000 archive materials

-The yearly reports increment is about 600-700 new reports

Keywords: Digital Archive

1 Úvod

Činnosť ŠGÚDŠ je zameraná na riešenie úloh geologického výskumu a prieskumu, tvorbu a využívanie informačného systému v geológii, registráciu, zhromažďovanie, evidenciu a sprístupňovanie výsledkov geologických prác vykonávaných na území Slovenskej republiky.

Odbor Geofondu v rámci rezortu MŽP SR predstavuje moderné archívne pracovisko pre oblasť geológie. Súčasťou riešenia geologickej úlohy „Geologický informačný systém GeolS“ je zavedenie Digitálneho archívu Geofondu.

2 Archív Geofondu

Archív Štátneho geologického ústavu Dionýza Štúra predstavuje unikátny fond výsledkov geologických prác z územia Slovenskej republiky. Na štúdium je k dispozícii voľne prístupná študovňa, umožňuje bádateľom prístup k archívnym dokumentom a archívnym pomôckam na štúdium, zabezpečuje vyhotovovanie odpisov, výpisov a kópií. Vyhľadávanie správ sa realizuje pomocou katalogizačných lístkov. Na internete sa nachádzajú signálne údaje o záverečných správach. Tento zoznam bol raz za mesiac aktualizovaný. Pracovníci archívu na požiadanie pripravujú zo správ rôzne rešerše. V archíve Geofondu sa nachádzajú správy od roku 1987. Archív uchováva viac ako 87000 archívnych materiálov. Každý rok do archívu pribúda 600-700 nových správ.

3 Predchádzajúci systém

V predchádzajúcom evidenčnom systéme sa evidovali iba metaúdaje a nie správy samotné. Dáta boli ukladané v proprietárnom formáte a preto neboli prístupné iným aplikáciám. Systém nie je integrovaný s ostatnými systémami ŠGÚDŠ. Pred publikovaním metadát o správach na internete bolo potrebné ich exportovať a importovať do oddelenej databázy. Pri prístupe k databáze cez internet používateľ videl iba niektoré metadáta, nevidel najaktuálnejšie správy, nebolo možné zobrazit' alebo prehľadať texty správ a príloh. Medzi ďalšie nedostatky systému patrila jeho nestabilita, nejednoznačnosť pri evidencii niektorých metaúdajov a nepraktická tlač katalogizačných lístkov. Za samotnými správami bolo potrebné vycestovať do Bratislavy. Vyhľadávanie konkrétneho textu prebiehalo manuálne v bádateľni.

4 Ciele projektu

V roku 2005 bol schválený ministrom životného prostredia SR projekt geologickej úlohy Geologický informačný systém GeolS. Jedným z cieľov projektu je vytvorenie digitálneho archívu ŠGÚDŠ. Za týmto účelom bol rozpracovaný systém, ktorý umožní:

- stabilnú a jednotnú evidenciu, ktorá bude dostupná prostredníctvom intranetu aj internetu - evidenciu správ - metadáta, skenované súbory, fulltextový index
- prepojenie na GIS a iné systémy ŠGÚDŠ
- efektívne vyhľadávanie
 - o všetky dostupné metadáta
 - o fulltext
 - o zložitejšie vyhľadávacie podmienky
- vytvorenie rešerše cez internet bez nutnosti komunikovať s pracovníkmi Geofondu
- sledovanie návštevnosti
- praktickú tlač katalogizačných lístkov
- konsolidáciu používaných číselníkov
- verzionovanie správ, ak je potrebná oprava metadát alebo skenovaného súboru

Vzhľadom na prístup cez internet musí systém spĺňať vysoké nároky na bezpečnosť a riadenie prístupu k správam (detaily správ môžu vidieť iba zaregistrovaní používatelia).

5 Realizácia projektu

V rámci spolupráce s firmou YMS, a.s. bola realizovaná analýza a návrh systému tak, aby spĺňal všetky vyššie uvedené požiadavky. Nasadené programové vybavenie zahŕňa IBM Content Manager, Oracle, Java aplikačný server Tomcat, webovú aplikáciu WebCM, ktorá sprístupňuje archív jeho používateľom. Počas riešenia projektu prebehla konfigurácia systému, úprava aplikácie pre potreby ŠGÚDŠ, import dát zo starého systému. Počas importu boli vytvorené číselníky a prebehla ich konsolidácia. Bola realizovaná oprava dát a import metaúdajov cca 80 000 správ.

Pre on-line prístup k bibliografickým údajom a informáciám so záverečných správ bol 1. apríla 2008 sprístupnený verejnosti Digitálny archív Geofondu. Do digitálneho archívu môžete vstúpiť anonymne, alebo ako registrovaný bádateľ. Anonymný bádateľ má možnosť iba vyhľadávania v rozšírených bibliografických údajoch a registrovaný má navyše umožnené fulltextové vyhľadávanie a otváranie pdf súborov naskenovaných textových častí správ.

6 Perspektíva

Pre uloženie správ bol vybraný server, kde je možné uložiť veľký objem dát. Predpokladá sa, že časom budú digitalizované všetky správy. V budúcnosti bude možné týmto spôsobom sprístupniť aj ďalšie archívne fondy ŠGÚDŠ. Použité technológie umožňujú rozšíriť použitie systému aj mimo archív. Systém slúži na evidenciu a prácu s dokumentmi akéhokoľvek druhu. Dokumenty je možné evidovať ako registratúrne záznamy a viesť tak správu registratúry.

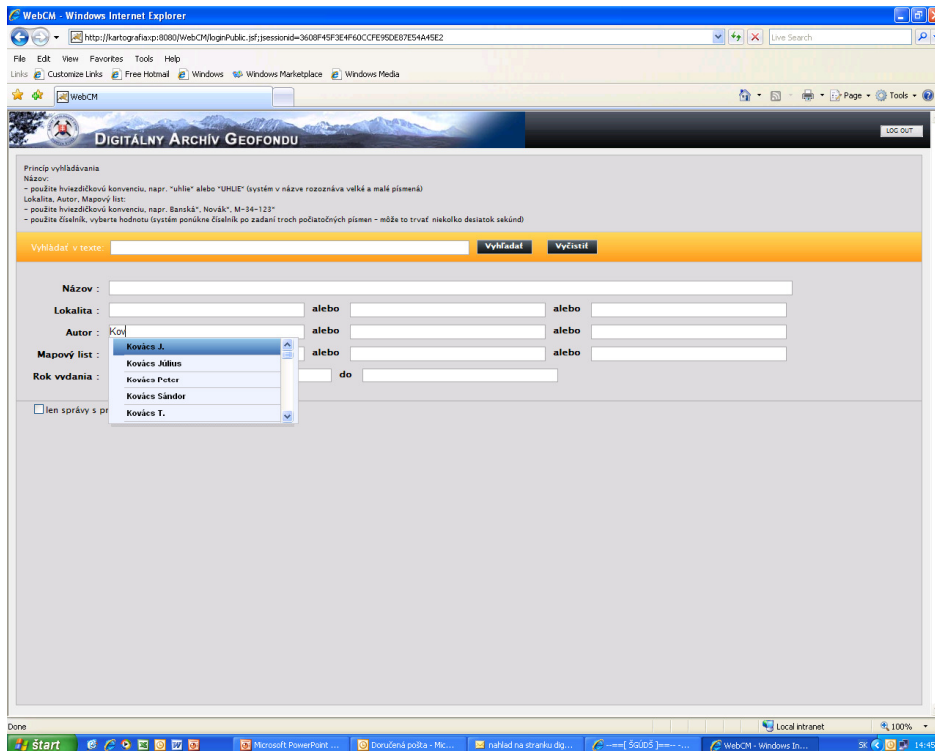
Ak majú s dokumentom pracovať postupne viacerí ľudia alebo systémy, je možné dokument zaradiť do pracovného procesu. Príkladom takéhoto procesu môže byť samotná digitalizácia správ, kedy je potrebné správy vybrať z archívu, skenovať, spracovať OCR, popísať pomocou metadát, vložiť do CM systému a vrátiť do archívu.



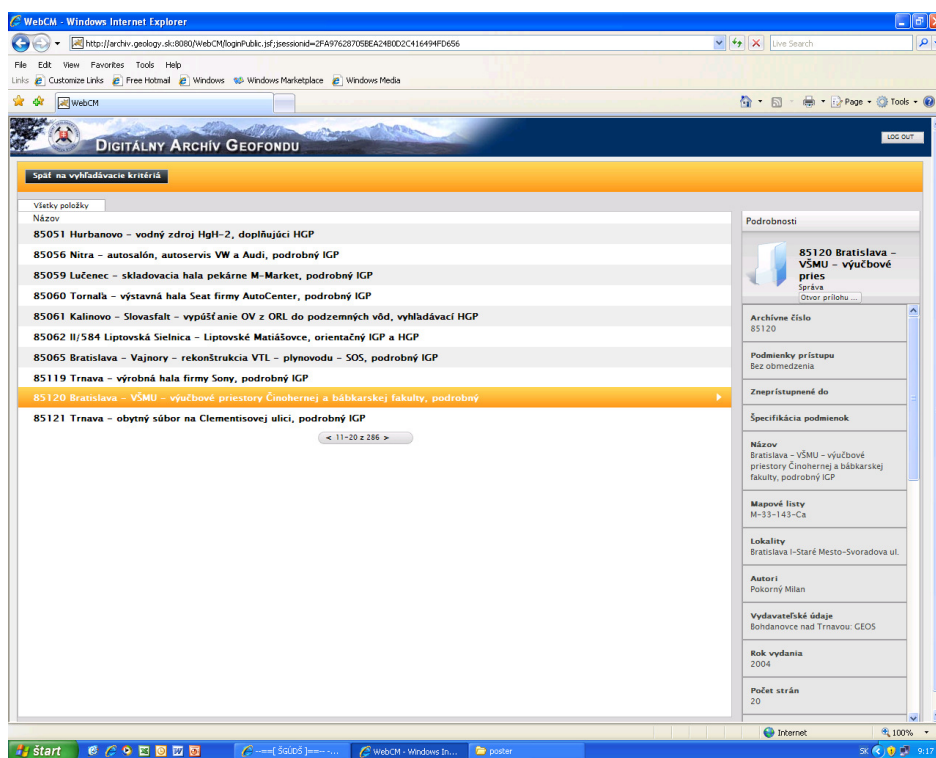
Obr. č. 1. ~ 87 000 správ, čo predstavuje cca 1,5 km



Obr. č. 2. Vstupná brána do Digitálneho archívu Geofondu na www.geology.sk



Obr. č. 3. Možnosti kombinovaného vyhľadávania v metainformáciách



Obr. č. 4. Výsledok vyhľadávania s možnosťou otvorenia záverečnej správy