

# Dilemata vědeckýchází dat

Jaroslav Šilhánek

Vysoká škola chemicko-technologická v Praze

[silhanek@vscht.cz](mailto:silhanek@vscht.cz)

INFORUM 2006: 12. konference o profesionálních informačních zdrojích  
Praha, 23. – 25.5.2006-04-12

**Abstrakt:** Necht' si pod označením „vědecká báze dat“ představujeme cokoliv, vždy je na jedné straně tvůrce takové báze dat (producent) a na straně druhé její uživatel, tedy vědec. A na obou stranách této symbiózy zákonitě dochází k situaci, kdy je nutné dělat více či méně závažná a někdy i protichůdná rozhodnutí, která pak zpravidla na velmi dlouhou dobu ovlivňují její fungování. Rozhodování o tom, jakým způsobem bude ta či ona vědecká báze dat vytvářena a následně využívána, je vždy podmiňováno technologiemi, které jsou v dané době dostupné a je tudíž evidentní, že nástup počítačů sehrál ve vývoji tohoto vztahu klíčovou úlohu a otevřel zcela nové možnosti. Přesto, že potenciál dnešních informačních technologií je zcela mimořádný a zdá se, že umožňuje prakticky vše myslitelné, stále zůstává řada otevřených otázek, které je při tvorbě každé vědecké báze dat řešit, a které pak logicky ovlivňují její využívání. K tomu přistupuje stále rychleji narůstající celkový objem vědeckých informací, který představuje dnes nekritičtější aspekt jakéhokoliv rozhodování na obou stranách a i celkového dalšího vývoje. Referát se pokusí shrnout nejčastější dilemata související jak s tvorbou, tak hlavně s aplikací vědeckých zdrojů informací při tvůrčí práci i s využitím konkrétních praktických zkušeností autora z obou stran problému.

## Vymezení pojmu „vědecká báze dat“

Před tím, než budeme hovořit a různých „dilematech“, aspektech, přednostech či nevýhodách a dalších problémech vědeckýchází dat, bude nepochybně účelné blíže vymežit tento pojem, respektive jeho obsah. Vyjdeme-li z prostého faktu, že nové poznatky, výsledky bádání, experimentů, pozorování nebo výpočtů jsou po staletí nejčastěji zveřejňovány písemnou formou článků ve vědeckých časopisech nebo v rozsáhlejší podobně knižních děl, pak nutně potřebujeme nástroj, který bude takovou produkci sledovat, evidovat a konečně archivovat, protože už dávno bylo zcela nemyslitelné, aby se taková produkce jen hromadila na regálech knihoven. Zatímco knižní produkce byla a celkem stále je zvládnutelná standardními katalogy knihoven, časopisecká produkce si vyžádala vznik samostatných nástrojů sumarizujících či průběžně sledujících a formou anotací registrujících zdrojů existujících buď v knižní podobě nebo v daleko efektivnější časopisecké podobě tzv. „referátových časopisů“. Producenti těchto referátových časopisů, jejichž historie jde hluboko do 19. století, se pochopitelně velmi brzy zajímali o možnosti využití výpočetní techniky a tak již od konce šedesátých let minulého století se setkáváme s jejich digitálními formami a také názvem „báze dat“, v zájmu přesnosti pak tedy „vědecké báze dat“. Některé takové báze dat teprve v tomto období vznikly

a nemají tedy svou tištěnou verzi a jiné opět tento přechod nezvládly a tudíž ukončily svou existenci.

V každém případě budiž konstatováno, že dále budeme hovořit o referátových zdrojích zpracovávajících převážně časopiseckou produkci vědeckých článků do podoby anotací, abstraktů nebo referátů, většinou tedy v konečné podobě bibliografických záznamů, ale v řadě případů pak i doplněných konkrétními daty, snad tedy s označením vědecké báze dat faktografické nebo bibliofaktografické. To už na dále diskutovaných problémech moc nemění.

Typickými příklady takovýchází bází dat jsou Chemical Abstracts, Compendex, Medline, Excerpta Medica, Inspec, Biosis, FSTA, GenBank, Metadex a řada dalších.

## **Dilema první**

Každý, kdo se rozhodne zpracovávat primární vědecké zdroje, ať již jako tištěný referátový časopis nebo dnes samozřejmě jako elektronickou bázi dat, musí vyřešit základní otázku, a to do jaké hloubky bude primární zdroj zpracovávat. Tedy zda se omezí na bibliografickou citaci s názvem, který indikuje obsah primárního zdroje, zda doplní odkaz abstraktem, a to převzatým z původního zdroje nebo nezávisle vytvořeným, a hlavně jak podrobně bude vlastní obsah indexovat v nejširším smyslu tohoto označení. Tato otázka je naprosto zásadní ze dvou důvodů:

- Jednou vytvořená koncepce zpracovávání primárních vědeckých zdrojů naprosto zásadním způsobem určuje možnosti využívání a může být v budoucnu jen obtížně měněna.
- V některých oborech, pro které je typické, že pracují s reálnými objekty a k nim přiřazenými informacemi, je nutné rozhodnout, zda tyto objekty budou součástí zpracovávání jako takové nebo nikoliv.

K první otázce dodejme, že ji lze chápat i v tom smyslu, že rozsah a hloubka zpracovávání primárních vědeckých dat představuje prakticky jedinou a poslední možnost, jak na výsledky tvůrčí vědecké činnosti upozornit, neboli jinak řečeno, nebude-li zachycen a označen nějaký důležitý aspekt výsledků bádání, je značně vysoká pravděpodobnost, že nebude nikdy nalezen a tudíž použit. Logickým důsledkem tohoto základního rozhodnutí jsou pak náklady na zpracovávání, protože čím podrobnější a odbornější má být takové intelektuální zpracovávání, tím bude náročnější a tudíž dražší. Jedná se tedy o dilema nepochybně klíčové a to čím dál tím více.

Ve druhém případě se jedná typicky o oblast chemie, biologie a řady dalších přírodovědných oborů, ale v zásadě i řady dalších oborů, kde přichází v úvahu důsledná a systematická excerpce konkrétních chemických sloučenin, organismů, přírodnin, ale i uměle vytvořených látek, objektů apod. V každém jednotlivém případě to pak může představovat problém, který svým rozsahem i vysoce překračuje dimenze původní báze dat. Příklad za všechny je opět chemie, kde báze dat chemických sloučenin je dnes svými více než 80 milióny záznamy individuálních látek vůbec největší známá báze dat a báze dat biologických fragmentů, ať již jako součást výše zmíněné báze nebo jako báze samostatné, tyto dimenze zcela určitě překročí.

## Dilema druhé

Jakmile se producent rozhodne o hloubce zpracovávání primárních zdrojů, musí hned vyřešit bezprostředně navazující otázku, a to **kdo bude zpracování provádět**. To je sice nepochybně naprosto zásadní otázka pro producenta, ale uživatel ji většinou nevnímá, i když se ho rovněž velmi zásadně týká. Principiálně máme dnes k dispozici dvě alternativy:

- Více či méně automatizované vyhledávání a shromažďování materiálů bez zásahu člověka, tedy bez intelektuální formy zpracovávání primárních zdrojů.
- Více či méně podrobné prostudování daného primárního zdroje, tedy jeho intelektuální zpracování člověkem.

Je evidentní, že první alternativa je atraktivní právě v současné době a každému asi bezprostředně napadne jeden nebo i více příkladů typu Google Science, Scirus a případně další. Pojem „automatizované“ zpracovávání ale měl (ale asi i stále má) svůj ekvivalent i v předpočítačové době. Bylo možné zpracovávat primární zdroje formou opisování názvů, bibliografické citace, klíčových slov autorů apod., tedy s žádnou nebo jen minimální účastí intelektuálního zásahu. V dnešní době ve výše uvedených příkladech úlohu zpracování přebírají algoritmy vyhledávacích strojů („search engines“), které sice už dnes a určitě tím více v budoucnu dosáhnou vysoké míry sofistikovanosti, ale zatím tato cesta představuje způsob naprosto zásadně odlišný od druhé možnosti. Budiž řečeno, že je to alternativa nepochybně už dnes určitě zajímavá a v žádném případě by neměla být z hlediska vědeckých bází dat programově opomíjená. Představuje ale svůj specifický přístup k vědeckým informacím, který nebude dále předmětem tohoto referátu. Lze jen uvést, že v současné době jsou nepochybně využívány některé nástroje automatizovaného vyhledávání vědeckých informací v plných textech primárních zdrojů jako pomocné nástroje při jejich intelektuálním zpracovávání a existují i takové nástroje komerčně dostupné a více či méně univerzálně použitelné. Existující nástroje umožňující fragmentaci dokumentu na specifické, snadněji dále zpracovatelné složky, jako je prostý text, experimentální části, tabulky, grafické složky apod., stejně tak existují zřejmě i úspěšně pracující aplikace pro automatizované vyhledávání chemických vzorců nebo názvů a dokonce vzájemné konverze grafického vzorce a názvu. Pokud je ale autoru známo i v těchto, nepochybně velmi sofistikovaných a úspěšných aplikacích, je nutné konečná intelektuální kontrola výsledků.

Alternativa intelektuálního zpracovávání primárních zdrojů představuje standardní způsob produkce referátových časopisů a zůstává stejná i pokud výsledným produktem je elektronická báze vědeckých informací. Tento aspekt možná zůstává při současné nabídce vědeckých bází dat v jejich konečné, uživatelsky příjemné formě pro uživatele skryt, ale v naprosté většině případů je alfou a omegou produkce vědeckých bází dat. Je logické, že v tomto bodě existuje velmi těsná souvislost s prvním dilematem, protože podle toho, jak podrobně a hluboce má být primární zdroj zpracováván, takovou úroveň musí mít ti, kdo tuto činnost budou dělat a samozřejmým důsledkem jsou pak finanční náklady na celou činnost.

Dokonce se zdá, že závažnost tohoto aspektu je čím dále tím větší. Tak, jak je vědecká činnost rozvíjena a řešený problém se dostává do větších podrobností a detailů, tak se prezentace výsledků stává méně srozumitelnou tomu, kdo sám nepracuje na velmi blízkém problému a nároky na profesionalitu a znalosti těch, kteří mají dnešní vědeckou produkci zpracovávat do vědeckých bází dat tak stále stoupají a tím stoupají i finanční náklady. Do budoucna představuje tento aspekt nepochybně velmi závažný problém, zatím se nabízí alespoň

částečné řešení ve zdokonalování a rozsáhlejším využívání různých pomocných nástrojů, o kterých byla řeč v předchozím odstavci.

Jistým, nicméně pro producenta velmi podstatným „podilematem“, je otázka, zda intelektuální zpracovávání primárních dokumentů bude svěřeno externím spolupracovníkům nebo si producent vybuduje a bude udržovat vlastní tým expertů. Tento aspekt konečný uživatel zpravidla vůbec nevnímá, první alternativa se přímo nabízí, zatímco druhá bude logicky podstatně finančně náročnější. Ovšem činnost externích spolupracovníků se obtížně sjednocuje a nemožnost pravidelných porad a konzultací jen velmi obtížně vede ke konsistentním výsledkům, což je daleko snazší, ale pochopitelně daleko dražší při týmu stálých pracovníků. Konečný uživatel to sice většinou nepozná, ale nakonec za to platí, protože dříve běžná spolupráce s externisty je dnes převážně nahrazena pracovníky „na plný úvazek“.

### **Dilema třetí**

Poměrně jednoduché může být rozhodnutí o oborovém záběru nebo o typu dokumentů, které budou nebo nebudou zpracovávány. S přihlédnutím k tomu, co již bylo výše o problematice vědeckých bází dat řečeno, je zřejmé, že se snahou pokrýt co nejširší oblast vědeckého bádání budou logicky narůstat výše uváděné problémy v případě, že uvažujeme variantu intelektuálního zpracovávání materiálu. Proto také vědecká báze dat pokrývající celou oblast vědy a techniky *de facto* neexistuje. Nejbližše se tomuto označení blíží Web of Science, tedy původní referátový časopis Science Citation Index, což je ale produkt, který není v pravém slova smyslu výsledkem intelektuálního zpracování dokumentů, protože sestává z bibliografické citace, klíčových slov autorů a „automaticky“ přebíraných citací jako alternativa k intelektuálně produkovanému věcnému popisu. Výběrovým kritériem pak není vědecký obor, ale prokazatelný aspekt ohlasu informačního zdroje, ve kterém jsou primární informace zveřejňovány.

Většina ostatních vědeckých bází dat je proto více či méně zřetelně oborově ohraničena, a i to, jako např. v chemických oborech, vede k produktu, který je označován jako vůbec největší elektronická báze dat, a to nejenom vědecká. Problematičnost šíře záběru ilustruje i nejnovější přírůstek do rodiny vědeckých bází dat, báze dat Scopus, která, přes svou nepochybně velkou šíři ve zpracování počtu vědeckých periodik, má poměrně značně nerovnoměrné oborové zastoupení, i když např. humanitní obory programově opomíjí. Zájem uživatelů je v tomto ohledu nejednotný, na jedné straně větší až hodně velká oborová šíře představuje možnost zachytit velmi cenné informace z hraničních oblastí jednotlivých vědních disciplín, na druhé straně jsou výsledky hledání často zahlceny z hlediska konkrétního zájmu balastem naprosto nerelevantních informací. Univerzálně se zdá, že vývoj půjde spíše směrem k velkým bázím dat, protože sledování řady navzájem se překrývajících zdrojů jen s tím účelem, aby náhodou nic neuniklo, je časově náročné a i přes dnes už poměrně velkou podobnost formálních nástrojů vyhledávání u jednotlivých zdrojů, stále existují a budou existovat větší či menší rozdíly komplikující práci.

Velmi podstatné je ale rozhodnutí, zda daná vědecká báze dat bude či nebude zpracovávat patentové zdroje. Že je zpracovávání patentů velmi obtížné, je všeobecně známé, v poslední době je to ilustrováno opět na příkladu báze Scopus, která sice patenty zahrnuje, ale z řady důvodů nejsou výsledky hledání patentů tak přehledné, jako u časopiseckých zdrojů. Není proto překvapující, že se jedná o doménu komerčního podnikání a produkty jako jsou World

Patent Index firmy Thompson – Derwent nebo PatentWeb firmy Micropatents jsou stále vysoce využívány, a to přes jejich značnou cenu. Skutečným problémem je pak fakt, že nezisková sféra má jen omezený přístup ke vědeckým informacím obsažených v patentových spisech.

### **Dilema čtvrté**

Zatím jsme se zabývali problémy, které jsou spojeny s koncepcí dané vědecké báze dat, tedy s návrhem jejího obsahu a zpracováváním materiálů. V okamžiku, kdy máme tyto otázky vyřešené a je nám tedy jasné co a jak budeme zpracovávat do vědecké báze dat, musíme rozhodnout další zásadní otázku, a to jak bude zpřístupňována. To je samozřejmě spojeno s ještě důležitější otázkou, a to jakým způsobem získáme zpět vložené investice, resp. jak velký bude zisk v případě, že produkce vědecké báze dat je založena na podnikatelském záměru. Samozřejmě tato otázka stojí ve skutečnosti na samém počátku jakýchkoliv úvah o produkci vědecké báze dat a podle toho, jak byla zodpovězena, budeme řešit i dilema o způsobu zpřístupňování a je rovněž samozřejmé, že bude-li vědecká báze dat koncipována jako služba financovaná z veřejných prostředků, viz příklad báze dat Medline, budou rozhodnutí o způsobu zpřístupňování jiná, než v případě, že požadavek na návratnost nebo dokonce zisk bude kategorický.

Omezíme-li, jak už bylo v úvodu konstatováno, své úvahy jen na elektronické báze dat, pak máme dvě zásadní možnosti:

- Báze dat jako produkt instalovaný na vlastním počítači resp. v lokální síti.
- Přístup na vzdálený server, prakticky výhradně prostřednictvím Internetu.

V současné době je prvá alternativa, která v podstatě otevřela svět rozsáhlých vědeckýchází dat, dnes aktuální v případě relativně malých a specializovanýchází dat. Hlavní důvod pro toto řešení, tj. problémy se spolehlivým dlouhodobým vzdáleným přístupem jsou dnes prakticky vyřešeny a zbývají tak spíše negativní aspekty jako je nutnost lokální administrace zpravidla na dosti vysoké úrovni. Neboli řešení je dnes jednoznačně na straně vzdáleného přístupu.

Připomeňme ale aspekt, který reálně existuje, ale příliš se nezdůrazňuje, protože se s ním prakticky nedá nic dělat. Na rozdíl od tištěných verzí vědeckýchází dat, tedy referátových časopisů, které se logicky vázaly a stavěly v knihovnách na regály a prakticky vždy doplňovaly rejstříky a zůstávaly proto ve vlastnictví dané knihovny, jsou jejich elektronické verze a tím spíše vědecké báze dat existující pouze v elektronické podobě dostupné jen do té doby, pokud jsou hrazeny příslušné licenční poplatky, resp. předplatné. Jakmile je finanční úhrada ukončena, přístup i k dříve zaplaceným ročníkům je ztracen. Praxe Web of Science, kdy údajně jednou zaplacené ročníky zůstávají v majetku instituce, je spíše výjimkou. I když je pochopitelné, že vědecká báze dat musí být neustále doplňována a aktualizována, je to nepochybně značně negativní aspekt. Na druhou stranu je rovněž nutné přiznat, že i tištěné báze dat, tedy svázané ročníky referátových časopisů, relativně rychle ztrácejí svou hodnotu jakmile byl ukončen jejich odběr.

Jestliže tedy nejčastější situace je dnes taková, že se bude jednat o vědeckou bázi dat přístupnou prostřednictvím sítě na vzdálený server, zbývá otázka, za jakých podmínek bude umožněn přístup. Pomineme-li výjimečnou situaci, kdy je daná báze poskytována *de facto*

zcela zdarma, je zásadní otázkou způsob úhrady za přístup nebo licenční poplatky. V zásadě přichází v úvahu dvě možnosti:

- Úhrada za přístup vycházející z doby využívání, formulace dotazu a objemu nalezených a převzatých informací, tedy přímá relace mezi cenou a objemem informací. Tato forma je typická pro komerční a průmyslové instituce.
- Úhrada za přístup nebo licenční poplatky jsou stanoveny jako paušál formou ročního předplatného nebo licenčního poplatku. Toto je prakticky jediná forma přicházející v úvahu pro univerzitní prostředí.

Jsou samozřejmě možná řešení kombinující obě možnosti, např. předplacení určitého objemu dat nebo i času, jsou ale spíše blíže první variantě. Současně je ale nutné si velmi důrazně uvědomit, že první alternativa je současně řešením zcela univerzálním a umožňuje přístup k vědeckým bázím dat bez jakéhokoliv dalšího omezení. Zde je celkem jednoduše řešitelné dilema na straně uživatelů, kdy je možné a mělo by být proveditelné porovnání nákladů na provedení rešerše s úhradou za spotřebovaný čas a objem informací a frekvenci využívání dané vědecké báze dat za licencované období, tedy nejčastěji jeden rok. Jestliže předplatné za vstup do nějaké báze dat stojí cca 300.000 Kč ročně a nějaká instituce jej využije v průměru 1x týdně, jsou náklady na paušální přístup s největší pravděpodobností vyšší než úhrada za každý konkrétní vstup. Za takto zdánlivě jednoznačným řešením se ale skrývá řada dalších aspektů, z nichž nejzávažnější je otázka zvládnutí a hlavně udržování jisté pracovní rutiny, která je v případě práce pod tlakem bezprostředního utrácení peněz dosti důležitá.

Přístup za základě paušální roční platby je nepochybně velmi výhodný a efektivní, ale producent v tomto případě musí vycházet z určité minimální výše ročních licenčních poplatků, které jsou logicky pro řadu menších institucí a pochopitelně jednotlivců mimo jejich možnosti. A protože současně ve větší či menší míře je okruh legálních uživatelů licenčními smlouvami omezován, a to často jednoznačně jen na pracovníky a v případě akademického prostředí také na studenty dané univerzity, dochází k velmi závažnému a bohužel velmi rozšířenému dilematu:

Vědecký pracovník, který není zaměstnancem instituce předplácející nebo udržující licenční přístup k té či oné vědecké bázi dat, nemá k tomuto informačnímu zdroji prakticky žádný přístup. A protože úhrada přístupu pro výdělečnou sféru je univerzálně vyšší, ti její tvůrčí pracovníci, kteří nepracují v několika málo největších světových firmách udržujících výzkumná oddělení, jsou v tomtéž postavení, kdy nemají možnost celou řadu současných vědeckých bází dat využívat.

Producenti jsou si této situace pravděpodobně dobře vědomi, ovšem jsou celkem pochopitelně vedeni snahou udržet přiměřenou vazbu mezi náklady na tvorbu vědecké báze dat a návratností dané investice nebo i zisku a proto není překvapující, že postupují značně obezřetně směrem k více benevolentnímu otevírání přístupu. V každém případě představuje tato situace, kdy jak menší univerzity nebo katedry na jednotlivých víceoborových fakultách, tak i střední průmysl, mají velmi omezený nebo žádný přístup k vědeckým bázím dat, jeden z největších problémů jejich širšího a tudíž efektivního využívání.

## Dilema páté

Výčet důležitých rozhodnutí, které je nutné udělat jak v situaci, kdy vědeckou bázi dat chceme vytvořit, tak i když ji chceme využívat, může být ještě daleko početnější, ale výše uvedené aspekty pokryly alespoň ty hlavní a klíčové body. Závěrem se soustředíme na určitě velmi závažnou otázku, a to jsou perspektivy dalšího vývoje a vůbec budoucnosti vědeckýchází dat. Jinými slovy na otázku, zda v éře elektronických zdrojů a počítačových sítí budou mít vědecké báze dat, které, jak jsme zdůrazňovali v úvodu, jsou většinou digitalizovanými verzemi tištěných referátových časopisů, své opodstatnění a místo.

Zatím jsme nezdůrazňovali možná nejzávažnější dilema související s vědeckými bázemi dat, a to je neustále narůstající objem primárních zdrojů a i narůstající rychlost jejich tvorby, která mimo jiné souvisí jak s objemem finančních prostředků vkládaných do vědy a výzkumu, tak i se současnými počítačovými technologiemi. Producenti vědeckýchází dat založených na intelektuálním zpracování musí dnes zvládat dvakrát tolik materiálu než před 10 - 15 lety, a to při nákladech, které nemohou růst stejným tempem, protože cena produktu by byla pro většinu uživatelů nepřijatelná.

Kromě samotného faktu stále narůstajícího objemu primárních vědeckých informací je zde možná ještě významnější a pro uživatele, který nemá vlastní praktické zkušenosti se zpracováním dat do vědeckýchází prakticky „neviditelný“ problém narůstající složitosti a náročnosti této činnosti. Je nutné si uvědomit, že vědecká práce v souladu se svým principem a motivací řeší stále detailnější problémy, které jsou značně vzdáleny od základních poznatků té či oné vědecké disciplíny a ten, kdo má takovou práci zpracovat do podoby abstraktu a případně excerpovat ten či onen konkrétní údaj, musí mít přibližně stejně hluboké znalosti, aby dané práci vůbec porozuměl. K tomu přistupuje i zájem vydavatelů udržet rozsah článků v pokud možno co nejstručnější podobě, což vede k maximálně úsporné presentaci bez obsáhlejšího výkladu souvislostí, které zasvěcený čtenář stejně musí znát. Je pak jen logickým důsledkem, že požadavky na kvalifikaci abstraktorů či excerpторů nebo analytiků jsou velmi vysoké a tito lidé pak musí být i přiměřeně zaplacení, čímž opět stoupají náklady na tuto činnost. Autor této úvahy měl možnost vykonávat podobnou činnost před cca 15-20 lety, a to ještě při zpracování spíše starší produkce a znovu v současné době a na základě této zkušenosti může konstatovat, že tento problém je nepochybně jeden z nejzávažnějších.

Znovu se tedy otevírá otázka, zda počítačové technologie nenabídnou nástroj, který by zpracovával primární zdroje automaticky jako plné texty a vyhledávali bychom informace bez mezičlátku vědeckýchází dat. Stručně jsme se této otázce dotkli už při druhém dilematu a zde lze jen opakovat, že v současné době je to možnost ještě značně vzdálená. Kromě komplexity vlastní vědecké problematiky přistupuje i zatím prakticky neexistující standardizace publikačních forem, a to i mezi tituly jednoho vydavatelství, což často souvisí se specifickými problémy té či oné vědecké disciplíny. Kromě těchto faktů, na jejichž základě je existence vědeckýchází dat do budoucna jen těžko nahraditelná, přistupuje i skutečnost, že jejich původní klasická role prostředníka mezi souvisejícími primárními zdroji dostala v současné době zcela nový rozměr. Stále rozsáhlejší propojení digitálních verzí jak sekundárních, tak i primárních zdrojů vede k tomu, že sekundární vědecká báze dat je velice efektivním přepínačem mezi jednotlivými odkazy a umožňuje tak velmi rychle získat nejenom odkaz, ale přímo konečnou informaci i vše, co s ní souvisí.