

Univerzita Karlova v Praze  
Filozofická fakulta  
Ústav informačních studií a knihovnictví  
*Informační věda*

**Jan Pokorný**

**SOUČASNÉ MOŽNOSTI  
INTEGRACE INFORMAČNÍCH ZDROJŮ**

**CURRENT MEANS OF LIBRARY  
INFORMATION RESOURCES INTEGRATION**

**dizertační práce**

**2007**

vedoucí práce: Doc. PhDr. Rudolf Vlasák

**Anotace**

Práce ukazuje trendy a současné možnosti integrace informačních zdrojů knihoven, které se snaží zpřístupňovat své tradiční i elektronické fondy kompaktním a unifikovaným způsobem pro co nejširší uživatelskou základnu. Cílem těchto snah za pomoci moderních technologií je poskytovat online služby s využitím celého informačního portfolia, které má knihovna k dispozici, a to 24 hodin denně. Práce shrnuje zkušenosti a nejosvědčenější postupy budování informačních systémů pro integraci informačních zdrojů.

**Klíčová slova**

integrace informačních zdrojů, informační zdroje, informační systémy, projektování integračních systémů, architektura integračních systémů

**Abstract**

This work discusses future trends as well as current means of library information resources integration providing compact and unified access to both, traditional and electronic collections to broad range of users. These modern-technology integration efforts aim to provide 24 hours a day, 7 days a week, online services using the complete information portfolio available to the library. The paper summarizes first-hand experiences and well-proven practices of the development of information systems for information resources integration.

**Keywords**

integration of information resources, information resources, information systems, integration system planning, integration system architecture

**Čestné prohlášení**

Prohlašuji, že jsem disertační práci vykonal samostatně s využitím uvedených pramenů a literatury.

## OBSAH

<b>PŘEDMLUVA.....</b>	<b>8</b>
<b>ÚVOD.....</b>	<b>9</b>
<b>1 CO JE INTEGRACE INFORMAČNÍCH ZDROJŮ.....</b>	<b>10</b>
1.1 CHARAKTERISTIKA INFORMAČNÍCH ZDROJŮ .....	10
1.1.1 Způsob záznamu informace.....	10
1.1.2 Typ obsažené informace .....	11
1.1.3 Omezení přístupu k informaci .....	12
1.1.4 Dostupnost informačního zdroje.....	12
1.1.5 Nabízené uživatelské rozhraní .....	13
1.1.6 Typ dotazovacího jazyka.....	13
1.1.7 Způsob ukládání a vyhledávání dat .....	14
1.1.8 Prezentovaný formát informací .....	15
1.2 PŘÍČINY VZNIKU SNAH O INTEGRACI INFORMAČNÍCH ZDROJŮ .....	15
1.2.1 Nalezení relevantních informačních zdrojů pro vyhledávání .....	17
1.2.2 Vyhledávání informací o primárních dokumentech v relevantních zdrojích .....	17
1.2.3 Zobrazení nebo dodání primárního dokumentu/digitálního objektu .....	18
1.3 CÍLE A PRINCIPY INTEGRACE INFORMAČNÍCH ZDROJŮ .....	18
1.4 RIZIKA INTEGRACE .....	20
1.4.1 Vysoké náklady integrace .....	20
1.4.2 Větší složitost integračního systému .....	20
1.4.3 Závislost na externích zdrojích .....	20
1.4.4 Větší a rychlejší dopad případných výpadků integračního systému .....	20
1.4.5 Hrozba průměrnosti .....	20
1.4.6 Nedostatečné zaškolení uživatelů a malá propagace .....	21
1.5 SLOŽKY A ÚROVNĚ INTEGRACE INFORMAČNÍCH ZDROJŮ .....	21
1.5.1 Přidaná hodnota .....	21
1.5.2 Zaměření na uživatelské skupiny .....	22
1.5.3 Modely propojování informačních zdrojů .....	23
1.5.4 Úrovně integrace .....	24
1.5.5 Typické skupiny informačních zdrojů z hlediska integrace v knihovnách .....	25
1.6 PROJEKTOVÁNÍ IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ.....	26
1.6.1 Metodologie projektování IS .....	26
1.6.2 Projektové vrstvy.....	28
1.7 ETAPY PROJEKTOVÁNÍ A INTEGRACE .....	29
1.7.1 Etapy projektování a implementace integračního systému.....	29
1.7.2 Etapy integrace zdrojů do systému.....	29
1.8 SHRNUÍ .....	30

<b>2</b>	<b>ARCHITEKTURA IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ .....</b>	<b>31</b>
2.1	ZÁKLADNÍ PRINCIPY ARCHITEKTURY IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ .....	31
2.1.1	<i>Globální architektura IS pro integraci informačních zdrojů .....</i>	<i>32</i>
2.1.2	<i>Funkční architektura IS pro integraci informačních zdrojů .....</i>	<i>33</i>
2.1.3	<i>Procesní architektura IS pro integraci informačních zdrojů .....</i>	<i>34</i>
2.1.4	<i>Technologická architektura IS pro integraci informačních zdrojů .....</i>	<i>35</i>
2.1.5	<i>Softwarová architektura IS pro integraci informačních zdrojů .....</i>	<i>40</i>
2.1.6	<i>Licenční podmínky využívání informačních zdrojů .....</i>	<i>41</i>
2.1.7	<i>Zajištění dostupnosti zdrojů a jejich obsahu .....</i>	<i>42</i>
2.1.8	<i>Manuálně prováděné služby jako náhrada nedostatečné automatizace .....</i>	<i>43</i>
2.2	UŽIVATELSKÉ ROZHRANÍ .....	44
2.2.1	<i>Orientace uživatele v rozhraní IS .....</i>	<i>45</i>
2.2.2	<i>Možnosti navigace .....</i>	<i>45</i>
2.2.3	<i>Možnosti vyhledávání .....</i>	<i>45</i>
2.2.4	<i>Konsolidace výsledků .....</i>	<i>45</i>
2.3	AUTENTIKACE A AUTORIZACE .....	46
2.4	BEZPEČNOST .....	48
2.5	AKVIZICE INFORMAČNÍCH ZDROJŮ .....	49
2.6	POPIS A ORGANIZACE INFORMAČNÍCH ZDROJŮ .....	52
2.6.1	<i>Popis informačních zdrojů .....</i>	<i>52</i>
2.6.2	<i>Popis různých typů dokumentů a integrace .....</i>	<i>53</i>
2.6.3	<i>Možnosti popisu podle stupně digitalizace a uspořádanosti .....</i>	<i>54</i>
2.6.4	<i>Organizace informačních zdrojů .....</i>	<i>55</i>
2.7	PROBLEMATIKA VYHLEDÁVÁNÍ V INTEGRAČNÍCH SYSTÉMECH .....	56
2.7.1	<i>Vyhledávací proces v integračním systému .....</i>	<i>56</i>
2.7.2	<i>Dotazovací a selekční jazyky .....</i>	<i>61</i>
2.7.3	<i>Problematika vyhledávací masky a dalších rozhraní pro vstup dat .....</i>	<i>62</i>
2.7.4	<i>Jazyk dokumentu a problematika vícejazyčnosti při paralelním vyhledávání .....</i>	<i>63</i>
2.7.5	<i>Konverze znakových sad .....</i>	<i>63</i>
2.7.6	<i>Konverze a transformace dotazů s ohledem na selekční údaje a indexy .....</i>	<i>64</i>
2.7.7	<i>Zpracování a prezentace výsledků .....</i>	<i>65</i>
2.8	PROPOJENÍ INFORMAČNÍCH ZDROJŮ A SLUŽEB .....	68
2.8.1	<i>Služby postavené na vyhledaných objektech .....</i>	<i>68</i>
2.8.2	<i>Služby spojené s uživatelskou schránkou - personalizované služby .....</i>	<i>70</i>
2.8.3	<i>Služby spojené s dodáním primárních informací .....</i>	<i>70</i>
2.8.4	<i>Problematika plateb .....</i>	<i>72</i>
2.9	VÝZNAM STATISTIK PRO INTEGRACI A VYUŽITÍ INFORMAČNÍCH ZDROJŮ .....	73
2.9.1	<i>Sledování IS z technických důvodů .....</i>	<i>73</i>
2.9.2	<i>Sledování chování uživatelů .....</i>	<i>73</i>

2.9.3	<i>Sledování využití zdrojů z akvizičních důvodů</i>	74
2.9.4	<i>Sledování využití zdrojů z ekonomických důvodů</i>	74
2.9.5	<i>Zpětná vazba pro další vývoj projektu</i>	74
2.10	PORTÁL JAKO ZASTŘEŠUJÍCÍ NÁSTROJ INTEGRACE	75
2.10.1	<i>Co je informační portál ve smyslu integrace informačních zdrojů</i>	75
2.10.2	<i>Co je portlet</i>	76
2.10.3	<i>Informační portál a možnost spolupodílet se na vytváření obsahu</i>	77
2.11	SHRNUTÍ	77
<b>3</b>	<b>TECHNICKÉ NÁSTROJE INTEGRACE INFORMAČNÍCH ZDROJŮ</b>	<b>78</b>
3.1	VÝZNAM TECHNICKÝCH STANDARDŮ	78
3.2	TECHNOLOGIE PRO VYHLEDÁVÁNÍ A KOMUNIKACI SE VZDÁLENÝMI INFORMAČNÍMI ZDROJI	79
3.2.1	<i>Z39.50</i>	79
3.2.2	<i>SRU/SRW</i>	81
3.2.3	<i>OAI-PMH</i>	82
3.3	TECHNOLOGIE NA ZAJIŠTĚNÍ OCHRANY AUTORSKÝCH PRÁV A LICENCÍ	82
3.3.1	<i>Označení původce dokumentu a vlastníka autorských práv</i>	83
3.3.2	<i>Řízení přístupu k obsahu dokumentu</i>	84
3.3.3	<i>Řízení práv k manipulaci s dokumentem</i>	84
3.3.4	<i>Řízení expirace platnosti a znehodnocení dokumentu</i>	85
3.3.5	<i>Zajištění kompaktního formátování dokumentu</i>	86
3.3.6	<i>Formát PDF pro komplexní řízení ochrany dokumentu</i>	86
3.4	TECHNOLOGIE VZDÁLENÉHO PŘÍSTUPU	87
3.4.1	<i>Proxy server</i>	88
3.4.2	<i>VPN</i>	88
3.4.3	<i>Terminálový přístup</i>	89
3.4.4	<i>Shibboleth</i>	89
3.4.5	<i>Technologie SSO</i>	90
3.5	TECHNOLOGIE PRO VYTVÁŘENÍ OBSAHU INFORMAČNÍCH ZDROJŮ	90
3.5.1	<i>Digitální repozitáře</i>	90
3.5.2	<i>Publikační systémy</i>	91
3.5.3	<i>Databáze</i>	91
3.6	SHRNUTÍ	91
<b>4</b>	<b>INTEGRACE INFORMAČNÍCH ZDROJŮ A ŘÍZENÍ JAKOSTI</b>	<b>93</b>
4.1	SYSTÉM ŘÍZENÍ JAKOSTI	93
4.2	ODPOVĚDNOST VEDENÍ	94
4.3	ŘÍZENÍ ZDROJŮ	94
4.4	REALIZACE SLUŽEB	94
4.5	MĚŘENÍ ANALÝZY A ZLEPŠOVÁNÍ	95
4.6	CERTIFIKACE	95

<b>5</b>	<b>MODELOVÝ PŘÍKLAD INTEGRACE INFORMAČNÍCH ZDROJŮ .....</b>	<b>96</b>
5.1	ÚVODNÍ ZPRÁVA – SOUČASNÁ SITUACE.....	96
5.2	VIZE .....	96
5.3	CÍL.....	97
5.4	ÚKOLY .....	97
5.5	POŽADAVKY NA FUNKCE .....	98
5.6	PŘIDANÁ HODNOTA PROJEKTU.....	98
5.6.1	<i>Přidaná hodnota pro uživatele .....</i>	<i>98</i>
5.6.2	<i>Přidaná hodnota pro zúčastněné knihovny.....</i>	<i>99</i>
5.7	KRITICKÉ FAKTORY .....	99
5.8	ŘEŠENÍ.....	99
<b>6</b>	<b>BEST PRACTICES.....</b>	<b>102</b>
6.1	PLÁNOVÁNÍ A ŘÍZENÍ INTEGRAČNÍHO PROJEKTU .....	103
6.1.1	<i>Zahájení projektu integrace informačních zdrojů.....</i>	<i>103</i>
6.1.2	<i>Zvážení rizik projektu .....</i>	<i>103</i>
6.1.3	<i>Kritické aspekty projektování integračního systému.....</i>	<i>104</i>
6.1.4	<i>Personální zajištění projektu.....</i>	<i>104</i>
6.1.5	<i>Finanční zajištění projektu .....</i>	<i>105</i>
6.1.6	<i>Zabezpečení a správa informačního systému .....</i>	<i>106</i>
6.1.7	<i>Havarijní plán.....</i>	<i>106</i>
6.1.8	<i>Zavedení integračního systému mezi uživatele .....</i>	<i>107</i>
6.1.9	<i>Propagace služeb integračního systému .....</i>	<i>107</i>
6.2	ARCHITEKTURA IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ .....	108
6.2.1	<i>Integrace klasického fondu.....</i>	<i>108</i>
6.2.2	<i>Integrace elektronických databází a archivů .....</i>	<i>109</i>
6.2.3	<i>Integrace volného prostoru internetu .....</i>	<i>109</i>
6.2.4	<i>Uživatelské rozhraní.....</i>	<i>110</i>
6.2.5	<i>Orientace na uživatele a spolupráce s dalšími službami.....</i>	<i>111</i>
6.2.6	<i>Využití všech komunikačních kanálů .....</i>	<i>111</i>
6.2.7	<i>Volba produktů a technologií .....</i>	<i>112</i>
6.3	AKVIZICE INFORMAČNÍCH ZDROJŮ DO INTEGRAČNÍHO SYSTÉMU.....	112
6.3.1	<i>Důraz na primární dokumenty .....</i>	<i>112</i>
6.3.2	<i>Důraz na elektronickou podobu dokumentů.....</i>	<i>113</i>
6.3.3	<i>Pokrytí.....</i>	<i>113</i>
6.3.4	<i>Volba optimálního typu akvizice.....</i>	<i>113</i>
6.3.5	<i>Garance důvěryhodnosti informačního zdroje .....</i>	<i>114</i>
6.4	POPIS INFORMAČNÍCH ZDROJŮ .....	115
6.4.1	<i>Potřeba jednoznačného, globálního a trvalého identifikátoru zdroje.....</i>	<i>115</i>
6.4.2	<i>Popis zdrojů využitelný běžným uživatelem.....</i>	<i>115</i>

6.4.3	<i>Předmětová klasifikace zdrojů</i> .....	116
6.5	VYHLEDÁVÁNÍ V INFORMAČNÍCH ZDROJÍCH .....	116
6.5.1	<i>Důraz na přímé vyhledávání primárních informací</i> .....	116
6.5.2	<i>Distribuce dotazů z integračního systému do dílčích IS</i> .....	117
6.5.3	<i>Uživatelsky přívětivé vyhledávání</i> .....	117
6.5.4	<i>Paralelní vyhledávání</i> .....	118
6.6	PREZENTACE DAT V INTEGRAČNÍM SYSTÉMU .....	118
6.6.1	<i>Metadata vyhledaných informačních objektů</i> .....	118
6.6.2	<i>Zobrazení výsledků a možnosti dalšího zpracování</i> .....	119
6.7	NEJČASTĚJŠÍ OMYLY PŘI INTEGRAČNÍCH SNAHÁCH .....	120
6.7.1	<i>Jeden produkt pokryje všechny požadavky na IS</i> .....	120
6.7.2	<i>Sjednotíme standardy a konfigurace všech dílčích informačních zdrojů</i> .....	120
6.7.3	<i>Všechny dokumenty budou v blízké době v elektronické podobě</i> .....	120
6.7.4	<i>Uživatelé nemusí být zapojeni do projektování</i> .....	120
6.7.5	<i>Finanční náročnost projektu je dána pořizovacími náklady</i> .....	121
6.7.6	<i>Při spuštění IS musí být realizovány všechny plánované funkce</i> .....	121
6.7.7	<i>Není třeba nasazovat systém řízení jakosti</i> .....	121
6.7.8	<i>S integračním systémem začnou uživatelé sami pracovat</i> .....	121
<b>7</b>	<b>ZÁVĚRY</b> .....	<b>122</b>
7.1	DOSAŽENÉ VÝSLEDKY .....	122
7.2	POUŽITÉ METODY .....	122
7.3	VLASTNÍ VÝZKUM .....	122
<b>8</b>	<b>SLOVNÍK ZKRATEK</b> .....	<b>125</b>
<b>9</b>	<b>LITERATURA</b> .....	<b>130</b>

## PŘEDMLUVA

Již několik let se v rámci pracovní skupiny zabývám vývojem a provozem technického řešení celonárodního integračního projektu knihoven Jednotná informační brána, jehož řešiteli je Národní knihovna ČR a Univerzita Karlova v Praze. Moje práce pro tento projekt mi přinášela a stále přináší celou řadu podnětů a výzvy k řešení. Právě ony mě vedly k napsání této práce, ve které se snažím poukázat, v co největší šíři a komplexnosti, na problematiku spojenou s integrací informačních zdrojů.

Chtěl bych v první řadě poděkovat PhDr. Bohdaně Stoklasové a RNDr. Pavlovi Krbčovi, CSc., vedoucím projektu Jednotné informační brány, za to, že mě k projektu přizvali a dali mi spoustu inspirace. Stejně tak děkuji Doc. PhDr. Rudolfu Vlasákovi za vedení mé dizertační práce a velkou podporu při mém studiu. Děkuji také Bc. Martinu Ledínskému za skvělou týmovou spolupráci při vývoji portálových řešení a provozu paralelního prohlíдача a Mgr. Michalovi Josífkovi za cenné připomínky a konzultace. V neposlední řadě také děkuji své manželce za podporu a trpělivost.

„Šílenost je to, když se snažíme řešit problém vždy tím jistým způsobem a přitom očekáváme jiné výsledky.“

Albert Einstein



## ÚVOD

Cílem této práce je definovat, analyzovat, stanovit a vyložit nejlepší zkušenosti a doporučení (tzv. best practices) pro budování informačních systémů pro integraci informačních zdrojů v oblasti knihovních služeb. Součástí analýz je porovnávání řady vlastností IS zpřístupňujících informační zdroje. Nejedná se o metodickou příručku v pravém slova smyslu, protože metodika se musí nutně lišit u konkrétních integračních projektů a konkrétních institucí na základě předprojektových analýz a specifických podmínek. Při zpracování byly využity zkušenosti z realizovaných projektů Jednotné informační brány CASLIN (Česká republika), KIS3G (Slovenská republika) a e-Government (Nový Zéland).

Integrace informačních zdrojů je typem služby, která jasně odpovídá na otázku budoucí funkce knihoven. Současný globální stav informačních zdrojů je charakteristický nebývalým chaosem a neustále klesající efektivitou a využitelností. Tato nepříjemná situace pro uživatele je ohromnou výzvou pro knihovny, které v elektronickém věku hledají novou identitu a poslání. Knihovny mají historickou šanci sehrát úlohu integrátora informačních zdrojů v době, kdy je informace jedním z nejcennějších obchodních artiklů a největší strategickou výhodou.

První kapitola popisuje základní charakteristiku informačních zdrojů a hlavní příčiny potřeby jejich integrace. Ukazuje cíle, principy a rizika v integračním procesu, analyzuje dílčí složky a sumarizuje jednotlivé vývojové etapy z hlediska projektování integračního informačního systému. Slouží jako shrnutí a zároveň jako východisko pro další kapitoly.

Druhá kapitola podrobně popisuje jednotlivé komponenty systému pro integraci informačních zdrojů a ukazuje možné modely celkové architektury z pohledu uživatelského rozhraní, autentikace a bezpečnosti, vyhledávání, propojení s navazujícími službami, statistik a z pohledu vytváření komplexních portálových řešení.

Třetí kapitola představuje některé moderní technické nástroje pro integraci informačních zdrojů a ukáže příklady jejich nasazení v praxi.

Čtvrtá kapitola se zabývá významem řízení kvality při procesech projektování, implementace a provozování IS pro integraci informačních zdrojů a seznámí nás s možnostmi aplikace norem ISO 9001 do těchto procesů.

Pátá kapitola ukazuje na modelovém příkladu praktickou implementaci informačního systému pro integraci informačních zdrojů.

Šestá kapitola seznamuje s nejlepšími postupy, doporučeními a zkušenostmi při budování IS pro integraci informačních zdrojů.

## 1 CO JE INTEGRACE INFORMAČNÍCH ZDROJŮ

Cílem této kapitoly je shrnout všechna podstatná východiska pro zavádění řešení integrace informačních zdrojů v knihovnách a jim podobných informačních institucích, které se snaží nabízet kvalitní a efektivní informační služby svým uživatelům. Základní službou knihovny je dodávat uživatelům primární informace, které v daný okamžik uživatelé potřebují. Informace lze vyhledávat v informačních zdrojích, ale informačních zdrojů je nesmírně mnoho a jsou velice různorodé. Služby, které se snaží zajišťovat dodávání primárních informací, jsou různě zaměřené a roztržité. Integrace informačních zdrojů nabízí metody, jak různorodost a roztržité zdroje a služeb efektivně překonávat.

Abychom v dalších kapitolách přesně rozuměli významu projednávané problematiky, musíme na začátku stanovit přesný význam základních pojmů. Cílem těchto definic není snaha o vytvoření obecně platných termínů z oblasti informační vědy, ale pouze o vymezení významu pro účely této práce.

**Informačním zdrojem** (tradičně označovaným jako informační pramen) rozumíme v této práci obecně libovolný informační objekt, který obsahuje dostupné informace odpovídající informačním potřebám uživatele. Informační zdroj může být tištěný, zvukový, obrazový nebo elektronický, včetně zdrojů dostupných online.<sup>32</sup>

Informačními zdroji mohou být buď přímo **dokumenty** umístěné ve volném prostoru nebo **informační systémy** (v textu budeme používat zkratku IS), ať již klasické nebo elektronické, které mají za úkol shromažďovat, zpracovávat a transformovat informace a zprostředkovávat je uživateli nezávisle na jejich časovém a prostorovém rozptýlu.<sup>32</sup> Informační systémy mohou zpřístupňovat buď primární informační objekty nebo záznamy o nich.

V této práci budeme tedy za základní informační jednotky považovat informační objekty (primární informace) nebo záznamy o nich (sekundární informace), které ukládáme a vyhledáváme prostřednictvím informačních systémů (reprezentující primární, sekundární či terciální informace). Protože ale můžeme integrovat všechny typy informačních objektů, hovoříme o informačních objektech, záznamech i IS souhrnně jako o informačních zdrojích.

### 1.1 CHARAKTERISTIKA INFORMAČNÍCH ZDROJŮ

Máme-li hovořit o integraci informačních zdrojů, musíme nejprve analyzovat jejich klíčové vlastnosti, abychom věděli, jaké aspekty mohou ovlivnit proces integrace. Informační zdroje mohou nabývat nejrůznějších forem. Z hlediska praktického využití informačních zdrojů jsou nejvýznamnější tyto vlastnosti:

- dostupnost informačního zdroje (ovlivněná zejména způsobem záznamu informace, typem obsažené informace a omezením přístupu k informaci)
- nabízená uživatelská rozhraní
- typ dotazovacího jazyka
- způsob ukládání a vyhledávání dat
- prezentovaný formát informací na výstupu

Pro dnešního uživatele, který chce získávat informace prostřednictvím počítače nezávisle na čase a místě, jsou výše uvedené vlastnosti klíčové, protože rozhodují o tom, jestli uživatel získá informaci online ihned v reálném čase (elektronický primární dokument) nebo jestli získá pouze odkaz a vlastní informaci bude muset obstarat s prodlevou v elektronické formě (např. elektronické dodávání dokumentů) či fyzické formě (např. klasická výpůjčka či papírové kopie).

Podívejme se na tyto vlastnosti podrobněji.

#### 1.1.1 Způsob záznamu informace

Podle způsobu záznamu informací dělíme informační zdroje na dvě základní skupiny:

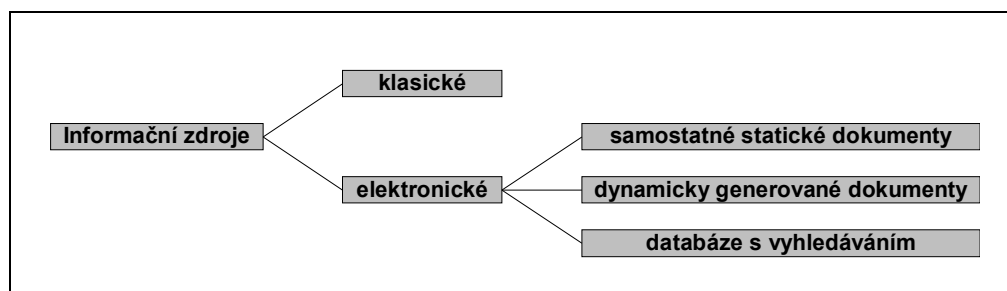
- informační zdroje v elektronické podobě (digitální)
- informační zdroje v jiné než elektronické podobě - označované též jako klasické či analogové

**Informace v elektronické podobě** jsou již jako elektronické vytvářeny nebo jsou do elektronické podoby převáděny (digitalizovány) z fyzických předloh. Elektronický dokument může sdílet velká skupina uživatelů ve stejný okamžik nezávisle na místě, podle momentálních potřeb uživatelů, na rozdíl od neelektronických (např. tištěných) dokumentů, které jsou fyzicky omezeny počtem exemplářů, těžkopádným způsobem přepravy apod. Výhodou elektronických informací je možnost je efektivně hromadně zpracovávat a transformovat do jiných podob. Význam elektronických informací rapidně stoupl rozšířením online služeb. Elektronické informační zdroje mohou být dostupné:

- jako samostatné statické dokumenty pomocí souborových, dokumentačních nebo publikačních systémů
- jako dynamicky generované dokumenty, jejichž obsah se mění
- formou klasických databází s vyhledávacím rozhraním

Významná část elektronických zdrojů je zpřístupněna prostřednictvím internetu v ohromném a nepřehledném kyberprostoru. Dynamicky generované dokumenty a databázové systémy tvoří část internetu, která se velmi obtížně indexuje a prohledává - konvenční vyhledávací nástroje ji pro proměnlivost obsahu a umístění často nedokáží postihnout. Tato část internetu se označuje jako hluboký nebo neviditelný web. Uživatelé jsou díky selhání indexovacích služeb o mnoho významných informačních zdrojů ochuzeni, protože se nedozvědí o jejich existenci. Důležitým jevem také je, že sice ohromné množství nových dokumentů na internetu každoročně vzniká, ale zároveň významná část obsahu internetu nenávratně mizí (až 40% ročně).<sup>6</sup> Problém zajištění trvalé dostupnosti je proto kritický.

**Informační zdroje v jiné než elektronické podobě** (klasické) představují všechny tištěné dokumenty, analogové multimediální dokumenty (nahrávky, snímky) apod. Vznikaly historicky před příchodem elektronických technologií zpracování, ale vznikají i v elektronickém věku. Období nástupu elektronických technologií se v oblasti klasických dokumentů vyznačuje dvěma významnými rysy. Za prvé, příchod digitálních technologií měl a má značný vliv na přípravu klasických dokumentů, zvláště v oblasti kvality, rychlosti a usnadnění přípravy (např. DTP pro přípravu tištěných dokumentů nebo digitální filmová studia pro zpracování klasických filmů). To je jeden z důvodů, proč produkce klasických, zejm. tištěných dokumentů neklesá, ale naopak vzrůstá. Klasické dokumenty tak nadále tvoří nedílnou součást celkového informačního portfolia. Druhým rysem je souběžné publikování stejných informací v elektronické i klasické podobě (souběžné vydávání časopisů v papírové i elektronické podobě atd.). Typické pro současnost také je, že sekundární zpracování klasických dokumentů se provádí již převážně elektronicky.



Obr. 1 - Základní typy informačních zdrojů podle způsobu publikování

### 1.1.2 Typ obsažené informace

Informační zdroje mohou nabízet:

- primární informace - vlastní informace (např. plný text, nahrávka, digitální objekt)
- sekundární informace - odkazy na primární informace (např. bibliografické záznamy)
- terciální informace - odkazy na sekundární informace (např. přehled databází)
- hybridní informace - kombinace primárních, sekundárních, popř. terciálních informací (např. bibliografické záznamy s referáty)

**Primární informace** představuje vlastní údaj o určitém aspektu reality (objekt, událost, idea atd.).<sup>32</sup> Umožňuje tedy uživateli získat okamžitou relevantní odpověď na daný dotaz a tak uspokojit svoji informační potřebu. Primární informace je prezentována formou plného textu

dokumentu, strukturovaného záznamu nebo multimediálního objektu bez ohledu na to, je-li elektronický nebo klasický.

**Sekundární informace** naopak obsahuje údaje o jiné informaci<sup>32</sup>, slouží tedy jako odkaz na primární informaci. Ve vztahu k této informaci má identifikační, popisnou a/nebo vyhledávací funkci. Koncovým uživatelům slouží sekundární informace jen jako pomůcka pro získání vlastních primárních informací. Pro další úroveň popisu a odkazování (terciální informace atd.) platí analogicky totéž, uživateli se o další fáze prodlouží proces získávání primárního dokumentu.

### 1.1.3 Omezení přístupu k informacím

Informační zdroje se odlišují **podmínkami přístupu**, tedy podmínkami, za jakých je producenti a provozovatelé zpřístupňují. Můžeme rozlišovat zejména informační zdroje:

- volné
- poskytované zdarma, ale na vyžádání
- placené (nutno zakoupit produkt nebo licenci)
- volné pouze pro skupinu (nutno se stát členem určité skupiny, např. členem)
- přístupné pouze prostřednictvím jiného subjektu (uživatel informace získá jako službu)
- nepřístupné (informace nejsou uživateli přístupné žádným legálním způsobem)

Druhým, stejně důležitým aspektem, který může výrazným způsobem omezit nebo zcela zabránit přístupu k informaci, je **vyhledatelnost**. O dokumentech v klasické i elektronické podobě se musí uživatel nejprve dozvědět, zjistit, že existují, a poté dohledat, kde se nacházejí, a získat k nim přístup. V tom mu mohou pomoci různé pomůcky, zejm. nástroje pracující se sekundárními informacemi. Při nabízení velkého množství informačních zdrojů je důležité zvolit efektivní způsob jejich organizace, aby se uživatel mohl ve velkém množství zdrojů zorientovat. Překážkou při získávání informací může být nejen absence informačních zdrojů, ale i jejich neorganizovanost a velké množství. S vyhledatelností také souvisí problematika neoficiálně publikovaných informací, publikování v rámci uzavřených skupin uživatelů (šedá literatura) nebo webových zdrojů neindexovaných běžnými indexovacími nástroji (tzv. neviditelný web).

### 1.1.4 Dostupnost informačního zdroje

U každého relevantního informačního zdroje je z pohledu uživatele nejdůležitější, jakým způsobem může uživatel zdroj využít. Mluvíme o dostupnosti informačního zdroje. Dostupnost je **kombinovaná vlastnost** a ovlivňuje ji trojice již zmíněných vlastností: způsob záznamu informace, typ obsažené informace a stupeň omezení přístupu. Moderní informační služby se snaží uživatelům nabízet v maximální míře online přístup ke zdrojům, aby všechny informace získali z monitoru svého počítače „tady a teď“. Informační zdroje mohou být nabízeny v následující souslednosti s klesajícím stupněm dostupnosti:

DOST.	PŘÍSTUP	FORMA	INFORMACE	PŘÍKLADY
↓	volný	elektronická	primární	plné texty na internetu, volné e-knihy, volně přístupné databáze, ...
	omezený	elektronická	primární	přístup k plným textům na základě členství, vlastní nákup elektronické knihy, ...
	volný	elektronická	sekundární	katalog knihovny na internetu, knihkupecké databáze na internetu, ...
	omezený	elektronická	sekundární	přístup ke komerční bibliogr. databázi na základě členství, vlastní nákup licence ke komerční bibliogr. databázi, ...
	volný	klasická	primární	bezplatná firemní dokumentace, prezenční výpůjčka deníku bez nutnosti členství
	omezený	klasická	primární	výpůjčka tištěné knihy na základě členství, vlastní nákup tištěné knihy, ...
	volný	klasická	sekundární	seznam doporučené literatury, bezplatný tištěný katalog vydavatele, ...
	omezený	klasická	sekundární	lístkový katalog knihovny na základě členství, vlastní nákup bibliografického soupisu, ...

U informačních zdrojů s omezeným přístupem má uživatel možnost využít členství v určitých organizacích, které pro své členy zajišťují přístupové licence (např. knihovny, půjčovny), nebo může nákup realizovat sám (přístup do databází nebo nákup samotného dokumentu).

IS by měly při integraci počítat se všemi zmíněnými typy informačních zdrojů, protože jsou v praxi významně zastoupeny. Uživatelům by je však měly nabízet sekvenčně s prioritou na primární elektronické dokumenty, aby uživatel nečekal na primární dokument déle, než je bezpodmínečně nutné.

### 1.1.5 Nabízené uživatelské rozhraní

Uživatelské rozhraní, jako rozhraní mezi uživatelem a IS, vytváří podle obecného pojetí pracovní prostředí, ve kterém uživatel komunikuje se systémem, zadává dotazy, případně jiná data, a kam se mu vrací výsledky, případně chybová hlášení nebo nápovědy.<sup>32</sup> Protože informační systém nemusí být pouze **elektronický**, musíme pro účely této práce mezi uživatelská rozhraní zahrnout i **fyzická** vyhledávací rozhraní, jako např. lístkový katalog. Můžeme tedy říci, že uživatelské rozhraní je v nejširším smyslu libovolné prostředí, které umožňuje uživateli pracovat s informačním zdrojem, zejm. vyhledávat a získávat informace. Můžeme je rozdělit do následujících skupin:

- tištěné seznamy a evidence, volné výběry dokumentů
- systémy lístkové evidence dokumentů (např. lístkový katalog)
- proprietární lokální systémy (např. GUI rozhraní databáze na CD-ROM)
- proprietární klienti online databází (např. DOS klient síťového AKS)
- webový prohlížeč jako klient online databáze (např. WWW rozhraní databáze Ebsco)

Formy uživatelských rozhraní jsou velice různorodé a ve větším počtu způsobují uživatelům **těžkosti při zvládnutí jejich ovládnutí**. Každý zdroj v dané skupině může nabývat různých podob a pojetí zpracování, navíc u elektronických rozhraní je výsledná podoba značně ovlivněna právě používanými technologiemi. Zvláště u aplikací vytvořených v 90. letech 20. století, kdy převládaly aplikace na CD-ROM a lokálně instalované systémy, je škála podob velice široká. Masovým rozšiřováním operačního systému MS Windows a webových prohlížečů dochází postupně ke sjednocování ovládacích prvků (formulářové prvky HTML, GUI rozhraní MS Windows), stále však existuje veliká diverzita rozhraní z hlediska designu, layoutu a pracovních postupů. Těžkosti uživatelů při práci s větším počtem heterogenních informačních zdrojů přetrvávají.

Z hlediska integrace je nutné zmínit ještě jiné rozhraní, než je uživatelské, a to **rozhraní systémové**, které slouží k výměně informací mezi dvěma IS. Pomocí systémového rozhraní, pokud jsou jimi IS vybaveny, lze interně bezešvě propojit systémy mezi sebou a plně je tak integrovat. Pokud systémové rozhraní chybí, lze k integraci za určitých podmínek použít i komunikační funkce uživatelského rozhraní, i když je určeno k jinému účelu. Nejrozšířenějšími typy systémových rozhraní pro informační zdroje používaných v knihovní praxi jsou technologie Z39.50, HTML/XML nebo obecné API (více informací o komunikačních rozhraních - viz kapitola 2.2).

### 1.1.6 Typ dotazovacího jazyka

Pracuje-li uživatel s elektronickými IS, používá k vyhledávání informací dotazovací jazyk. Z hlediska uživatele rozlišujeme jazyky umožňující výběr z **předdefinovaných** dotazů (např. menu), jazyky pro **listování** (browsing, navigaci) a **strukturované** jazyky s klíčovými slovy (např. SQL).<sup>32</sup> Dotazovací jazyky často obsahují také příkazy pro jiné operace, než je dotazování, a to nejčastěji editování obsahu. Příkladem může být již zmíněný SQL nebo Z39.50. Ve stadiu výzkumů je využití přirozeného jazyka ve funkci dotazovacího jazyka, v praxi se již objevují systémy, které podporují jazykovou flexi.

Zatímco u předdefinovaných dotazů a při listování obsahem jsou uživatelé v maximální míře obslouženi systémem a možnost, že by uživatel dotaz nesprávně zapsal, je minimalizována, u strukturovaných dotazovacích jazyků je tomu naopak. Uživatel vkládá buď celý dotaz nebo jeho parametry do vyhledávací masky ručně a způsob zápisu může významně ovlivnit výsledek vyhledávání. Z tohoto hlediska se implementace strukturovaných dotazovacích jazyků v IS dělí do následujících skupin:

- uživatel zapisuje celý dotaz ručně
- uživatel zapisuje vyhledávací podmínky, systém zkompletuje dotaz
- uživatel zapisuje vyhledávací podmínky, systém je zvaliduje a zkompletuje dotaz
- uživatel vybírá z nabídky podmínek, systém vytvoří vlastní dotaz

Lišit se tak nemusí pouze samotné dotazovací jazyky, ale i jejich implementace v IS. Implementace dotazovacího jazyka je výrazně ovlivněna uživatelským rozhraním. Moderní IS se snaží nabídnout taková rozhraní, která by možnost vzniku chyb způsobeném nesprávným zápisem dotazů eliminovala, ale takové snahy mohou **negativně ovlivnit** vyjadřovací schopnosti jazyka a možnosti vyhledávání. Přesto se ukazuje, že uživatelé dávají přednost **jednoduchosti** a teprve v případě neúspěšného hledání sahají po složitějších nástrojích.

### 1.1.7 Způsob ukládání a vyhledávání dat

Při integraci informačních zdrojů je důležité vědět, jakým způsobem jsou informace v podobě dat uloženy v IS a jak s nimi pracuje vyhledávací stroj. Fyzicky jsou všechna data uložena v souborech, protože soubor je základní prvek, se kterým je schopen pracovat operační systém. Soubor je logicky organizován jako posloupnost záznamů, které jsou ukládány do diskových bloků. Soubory mohou být organizovány sekvenčně, index-sekvenčně, indexově nebo s přímým přístupem. Organizace datového souboru spolu s návrhem vlastní datové struktury rozhodují o tom, podle jakých atributů (vyhledávacích bodů) a jak rychle může vyhledávací stroj uložené informace v souborech najít. Výrazně tak ovlivňují nejen rychlost, ale i možnosti vyhledávání.<sup>46</sup>

V prostých **sekvenčních souborech** jsou data za sebou ukládána do záznamů, buď zcela neuspořádaně, nebo uspořádaně podle určitého klíče. Uspořádání se však postupně narušuje s přibývajícím záznamy, které již nelze zařadit do uspořádané části a které se přechodně ukládají do speciální odkládací oblasti souboru (oblast přetečení). Je proto třeba provádět pravidelné údržby (reorganizace dat). Zpracování sekvenčních souborů je z výše uvedených typů souborů nejpomalejší.

► Vyhledávací stroj musí procházet postupně celý soubor, dokud nenarazí na hledaný údaj.

U **index-sekvenčních souborů** je sekvenční soubor seřazený podle primárního klíče doplněn o víceúrovňovou strukturu řídkého indexu (index je tvořen skokově pouze některými hodnotami klíče, např. každou desátou hodnotou). Pomocí indexu se v souboru přidávají nebo odebírají prvky, zefektivňuje se také vyhledávání. Nevýhodou je, že změna záznamů v sekvenčním souboru vyžaduje také změny v indexu. I u tohoto typu souboru je třeba provádět pravidelné údržby kvůli setřídění.

► Vyhledávací stroj může k primárnímu souboru přistupovat pomocí indexu, nemusí tedy postupně procházet celý primární sekvenční soubor. Může použít i jiné indexy obsahující další atributy.

**Indexový soubor** pracuje také se sekvenčním souborem, ale na rozdíl od předchozích dvou nemusí být setříděn, protože je použit hustý index, který na rozdíl od řídkého indexu obsahuje všechny hodnoty klíče. Indexů může být vytvořeno více pro různé vyhledávací klíče.

► Vyhledávací stroj používá k vyhledávání indexy, které jsou lépe organizovány a které jsou menší než primární sekvenční soubor, takže vyhledávání je výrazně rychlejší.

**Soubory s přímým přístupem** používají hašovací algoritmy k určení adresy bloku na disku, ve kterém je záznam uložen, na základě hodnoty primárního klíče.

► Vyhledávací stroj přistupuje přímo na místo, kde je záznam uložen, aniž by prohledával sekvenční nebo indexový seznam. Tento způsob organizace souboru je z hlediska vyhledávání nejrychlejší.

Data mohou být strukturovaná (např. MARC nebo XML), částečně strukturovaná (např. HTML) nebo zcela nestrukturovaná (např. prostý text), přičemž nejvíce vzniká dat nestrukturovaných. Nestrukturované dokumenty nelze vyhledávat podle polí, ale pouze plnotextově. Stupeň **strukturovanosti** dat na jednotlivé atributy (pole) je velice důležitý pro indexování a vyhledávání. Příkladem různých způsobů strukturování může být soubor s autorskými údaji. Jména autorů jsou tvořena křestním jménem a příjmením a lze je ukládat buď společně do jednoho pole nebo do dvou samostatných polí. Do společného pole je lze ukládat ve tvaru <jméno příjmení> nebo invertovaně ve tvaru <příjmení jméno>. Podobné možnosti existují i při budování indexu. Vyhledávací body rozhodují o tom, podle jakých údajů a jakým způsobem můžeme informační zdroj prohledávat. Pokud informační zdroj například podporuje vyhledávání jména autora pouze jako frázi v invertovaném tvaru, při opačném způsobu zápisu se nám nevrátí relevantní výsledek.

Z hlediska uložených dat můžeme rozlišovat několik **typů datových úložišť**:

- data uložená přímo v souborech operačního systému (struktura databáze je dána adresářovou strukturou)
- data uložená pomocí databázového stroje:
  - do strukturovaných souborů operačního systému
  - jiným způsobem přímo na záznamovém médiu
  - kombinace předchozích dvou metod (část dat v souborech OS, např. binární datové typy, část dat ve strukturovaných souborech, nebo přímo na záznamovém médiu)

Nebo můžeme rozlišovat typy datových úložišť z hlediska přístupu k úložišti:

- lokálně
- vzdáleně (dostupné síťovými prostředky)
- distribuovaně (na více místech)

Zatímco textové informace může vyhledávací stroj prohledávat přímo, u netextových informačních objektů (např. obrázků) je třeba evidovat doplňující údaje, podle kterých bude objekt vyhledatelný. Část z nich lze generovat automaticky z vlastností digitálního objektu (některá technická metadata), ostatní se musí vytvářet intelektuálně (popisná metadata).

### 1.1.8 **Prezentovaný formát informací**

Koncový uživatel dostává informace ve formátu, který mu nabídne na výstupu IS. Může mít na výběr více typů výstupních a zobrazovacích formátů, podle potřeby. Typické zobrazovací formáty jsou například:

- uživatelsky přívětivý výpis bibliografických záznamů s použitím popisů polí
- MARC
- formátovaný plný text
- tabulky
- netextové objekty

Původní data ukládaná v datových úložištích jsou často konvertována do podoby, které je pro člověka lépe čitelné, srozumitelnější a přehlednější. Např. údaje rozložené do relačních tabulek v relační databázi jsou prostřednictvím vyhledávacího stroje předávány aplikační vrstvě, která je převede do výměnného formátu MARC, uživateli jsou však informace prezentovány v tabulkovém, zjednodušeném zobrazení s popisem polí. V procesu vyhledávání může docházet k většímu počtu transformací formátů a protokolů (více informací o konverzích formátů - viz kapitola 2.7.6).

## 1.2 **PŘÍČINY VZNIKU SNAH O INTEGRACI INFORMAČNÍCH ZDROJŮ**

Počátky snah integrovat informační zdroje jsou spjaty se vznikem **systémové integrace** koncem osmdesátých let, kdy došlo k masovému nasazování osobních počítačů v organizacích. V té době se začaly projevovat problémy s nízkou efektivitou řady informačních projektů, zejm. vlivem vzniku mnoha proprietárních informačních podnikových systémů. První úvahy o systémové integraci proto přicházejí od lidí zabývajících se strategickým řízením podniku.<sup>49</sup>

V podnikovém pojetí se systémová integrace chápe jako komplex činností směřující k integraci jednotlivých komponent IS a služeb externích dodavatelů do jednoho produktu. Cílem systémové integrace je vytvoření a stálá údržba integrovaného IS, který optimálně využívá potenciálu dostupných informačních technologií k maximální podpoře podnikových cílů. Informační systém je přitom vytvářen spojováním různých zdrojů, tj. různých produktů a služeb.<sup>50</sup> Teorie integrace informačních zdrojů v oblasti služeb knihoven přejímá velkou část metodologie moderních směrů systémové integrace pro podnikové IS, má však i velké množství specifik.

Informační zdroje představují základní pomůcku pro uspokojování informačních potřeb uživatelů. Mají-li být informace obsažené v informačních zdrojích efektivně využitelné, musí být splněny následující podmínky:<sup>50</sup>

- uživatel musí být motivován získat a využít potřebnou informaci
- požadovaná informace musí existovat
- uživatel musí být schopen se dozvědět, že informace existuje
- uživatel ji musí být schopen získat
- uživatel ji musí být schopen porozumět a interpretovat

Podíváme-li se na tyto podmínky z pohledu knihoven, které mohou efektivitu využitelnosti informačních zdrojů ovlivnit, můžeme proces uspokojování informačních potřeb uživatelů rozdělit na 3 základní etapy:

- 1. nalezení relevantních informačních zdrojů pro vyhledávání**  
= dozvědět se, v čem hledat
- 2. vyhledávání informací o primárních dokumentech v relevantních zdrojích**  
= dozvědět se, co získat
- 3. zobrazení nebo dodání primárního dokumentu/digitálního objektu**  
= dozvědět se, jak to získat

Přirozeným vývojem zejm. v oblasti vědy a techniky však dochází k tomu, že informačních zdrojů neustále přibývá, a to takovým tempem a způsobem, že selhávají klasické způsoby využití založené na tradičním knihovnickém principu. Narůstá počet informačních zdrojů, o jejichž existenci nemají uživatelé ponětí a jsou velmi těžko vyhledatelné pomocí běžných vyhledávacích nástrojů. Informační zdroje, které uživatelé naleznou, nabízejí proprietární rozhraní a pracovní postupy. Na výstupu potom vrací informace prezentované v odlišných, proprietárních formátech a těžko se mezi sebou slučují. Proces uspokojování informačních potřeb uživatelů se proto neustále zpomaluje, komplikuje a bez nasazení vhodných řešení klesá jeho kvalita. Integrace informačních zdrojů nabízí metody, jak tuto situaci řešit.

**Snahou integrace informačních zdrojů** je nabídnout uživateli „bránu“, kterou bude přistupovat ke všem zdrojům a službám knihovny v bezpečném a pohodlném prostředí. K zajištění základní služby dodávání primárních informací dnes slouží řada izolovaných služeb, které je třeba konsolidovat. Integrační systém musí uživatele „navést“ k cíli, tj. pomoci mu určit, jakou službu a jaké informační zdroje vlastně potřebuje, a poté ho musí službou provázet až do fáze získání primárních informací. V dnešní době významných paradigmatických změn ve společnosti, na které má velký vliv informatizace a trend využívání nejrůznějších služeb na dálku, sílí tlak na knihovny, aby provozovaly takové knihovní služby, které v maximální míře omezí nutnost knihovnu fyzicky navštěvovat. Integrace zdrojů a služeb na jedno místo elektronicky dostupné na dálku je odpovědí na tuto potřebu.

Potřeba integrovat informační zdroje vzniká postupně a indikují ji **průvodní jevy**, a to zejména:

- uživatelé musí při vyhledávání ručně zpracovávat informace
- vyhledávací procesy vyžadují vstup uživatele
- objevuje se nutnost opětovného zadávání nebo přepisování dat
- dochází k častému výskytu chyb vyhledávání
- uživatelé nemohou zjistit stav/dostupnost požadovaných dokumentů

Problémy s využitím heterogenních zdrojů se zvyšují nárůstem jejich počtu. Ne každý uživatel vyžaduje ke své práci široké spektrum informací, potřeba integrace proto nikdy není u všech subjektů stejně silná.

Snaha integrovat informační zdroje nevychází pouze z potřeb koncových uživatelů, i když ty jsou samozřejmě nejdůležitější, ale i z fungování samotných informačních systémů a potřeb jejich administrátorů. Integrace zdrojů umožňuje informačním systémům mezi sebou lépe komunikovat a spolupracovat. V oblasti knihoven máme na mysli zejm. procesy výměny dat pro účely přebírání záznamů, sdílení katalogizace, datového sběru atd. Administrátorům integrace zjednodušuje práci při instalacích a konfiguracích dílčích informačních systémů na pracovní stanice a servery.

Technologie používané pro integraci informačních zdrojů se rychle mění vlivem technologií používaných u dílčích informačních zdrojů. Přelom 80. a 90. let 20. století byl typický masivním rozšířením lokálně instalovaných elektronických informačních zdrojů, zejm. na CD-ROM, v knihovnách a dalších organizacích, protože síťová konektivita na WAN byla rychlostně pomalá a finančně nákladná a nebyly vyvinuty potřebné síťové technologie. Začala se proto objevovat



integrační řešení pro zpřístupňování informačních zdrojů (včetně multimediálních) na CD-ROM a serverech v lokální síti.<sup>36</sup> Postupným rozšiřováním a zlevňováním internetu však vzrůstalo a stále vzrůstá využívání online zdrojů, resp. zdrojů dostupných síťovými protokoly typickými původně pro WAN (HTTP, Z39.50). Lokálně instalované informační zdroje, jejichž hlavní nevýhodou jsou omezené možnosti průběžných aktualizací a uzavřenost, pomalu ztrácejí na významu a mizí. V integračních projektech se však můžeme s lokálními proprietárními informačními zdroji stále setkávat, a proto je nutné mít k dispozici metody integrace i těchto zdrojů.

Důvody, proč integrovat informační zdroje, označujeme jako **integrační šance** a lze je shrnout do následujících bodů:

- přivedení uživatele blíže k plnému textu nebo digitálnímu objektu
- zjednodušení a sjednocení uživatelského rozhraní
- optimalizace chování informačních zdrojů
- zlepšení rešeršní výkonnosti
- vytvoření rámce pro průběžné zlepšování všech dílčích informačních zdrojů
- integrace na jedno místo elektronicky dostupné na dálku jako řešení pro hendikepované lidi a lidi fyzicky vzdálené, kteří těžko hledají zdroje lokálně

Integrační šance je třeba vždy zvažovat spolu s riziky integrace, o kterých podrobněji pojednává kapitola 1.4.

Podívejme se nyní podrobněji na jednotlivé etapy uspokojování informačních potřeb.

### 1.2.1 Nalezení relevantních informačních zdrojů pro vyhledávání

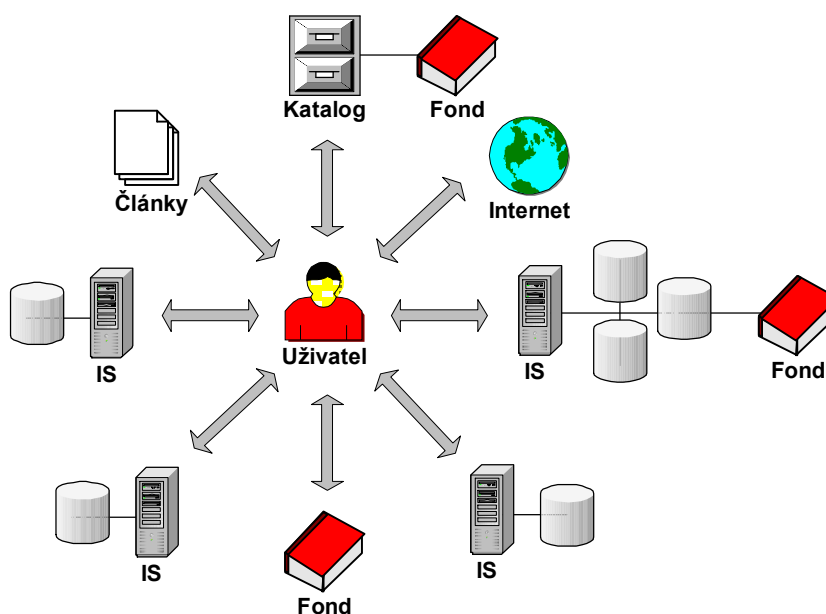
V současné informační praxi mají uživatelé k dispozici ohromné množství informačních zdrojů, ve kterých mohou vyhledávat a získávat informace (např. známý adresář Gale Directory of Online, Portable, and Internet Databases eviduje asi 15 600 databází od více než 4 000 producentů<sup>47</sup>). V této veliké chaotické množině musí uživatelé nejprve nalézt konkrétní informační zdroje, které jsou pro jejich informační potřebu relevantní. Existuje řada způsobů, jak tento proces řídit, a to jak u klasických informačních zdrojů, tak i u elektronických. Tyto metody většinou vycházejí z určité formy indexace, evidence a/nebo organizace do klasifikačních systémů, které dobře známe z oblasti klasického knihovnictví. Každý informační zdroj je možné opatřit jmenným a/nebo věcným popisem, jehož prvky potom využívá vyhledávací systém. Vznikají tak systémy terciálních informací. Ohromný nárůst počtu informačních zdrojů však znemožňuje popisovat všechny zdroje ručně a do popředí se dostávají otázky automatického zpracování, automatické indexace a automatické klasifikace.

### 1.2.2 Vyhledávání informací o primárních dokumentech v relevantních zdrojích

Jakmile uživatelé získají znalost o relevantních informačních zdrojích, které mohou použít k vyhledávání potřebných informací, objevuje se další problém, a to je vyhledávání samotných informací. Společným jmenovatelem informačních zdrojů, jak popisuje kapitola 1.1, je vysoká míra diverzity. **Diverzita informačních zdrojů** je způsobena jednak obecnými vlastnostmi informačních zdrojů a jednak technologicky. Informační systémy se vlivem rychle se měnících technologií a způsobů zpracování navzájem odlišují a jejich životní cyklus je delší než etapy uvádění nových technologií na trh. IS má tak tendenci zastarávat v prostředí, které se rychle vyvíjí.

Nastává situace, kdy má uživatel k dispozici bohaté portfolio informačních zdrojů, ale tyto zdroje nemůže plně využít, protože se **navzájem výrazně liší** a uživatel není schopen se všemi efektivně pracovat. Nabídka zdrojů se navíc velice rychle obměňuje, takže uživatel je nucen používat stále nové a nové nástroje, se kterými se musí vždy individuálně seznámit a naučit se je ovládat. Při velkém počtu zdrojů je to nemožné. Typicky se jedná o odlišnosti:

- rozhraní (proprietární rozložení, proprietární ovládací prvky)
- umístění (konkrétní adresa URL nebo umístění fyzické)
- síťové omezení v případě online zdrojů (porty a protokoly)
- autentikace (proprietární řízení přístupu, vlastní databáze uživatelů)
- dotazovací jazyk
- kódování znakových sad
- formáty dat a výstupní sestavy



Obr. 2 - Typické diverzní informační prostředí z pohledu uživatelů

Informace, které se v dnešní době rapidně mění a s časem jejich hodnota klesá, je nutné poskytovat „tady a teď“. Uživatel potřebuje mít ke všem požadovaným informacím přístup rychle a efektivně. Diverzita zdrojů je jednou z hlavních **překážek** uspokojování informačních potřeb uživatele. Je nutné hledat způsob, **jak diverzitu překonávat**. Řešení přináší moderní technologické prostředky umožňující **integrovat** heterogenní informační zdroje a uživateli nabízet **jednotné pracovní prostředí**.

### 1.2.3 Zobrazení nebo dodání primárního dokumentu/digitálního objektu

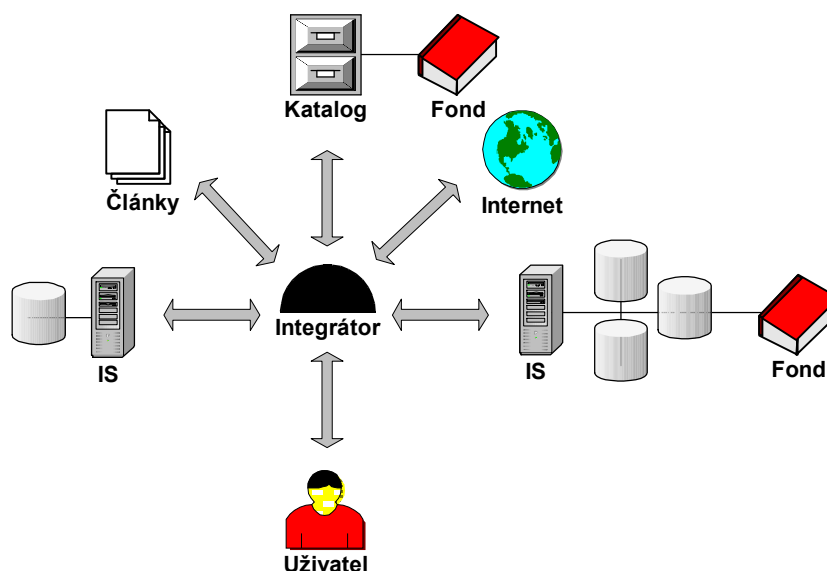
Jakmile uživatel vyhledá informace v informačních zdrojích, setkává se s posledním vážným problémem v řadě, a to získání vlastního primárního dokumentu či digitálního objektu. Jak jsme zmínili v předešlých kapitolách, dostupnost primárních dokumentů je různá. Některé dokumenty jsou k dispozici v elektronické podobě, některé v klasické podobě s možností jejich digitalizace v rámci navazujících služeb (např. elektronické dodávání dokumentů) a některé dokumenty v klasické podobě, které lze získat pouze fyzicky pomocí služeb typu MVS nebo běžné absenční/prezenční výpůjčky. Tyto různé formy služeb lze opět sjednocovat pomocí integračních metod.

## 1.3 CÍLE A PRINCIPY INTEGRACE INFORMAČNÍCH ZDROJŮ

Z dříve uvedených argumentů vyplývá, že primárním cílem integrace informačních zdrojů je nabízet relevantní informace jednoduchým a jednotným způsobem co možná nejširší skupině oprávněných uživatelů v co nejkratší době nezávisle na prostorovém rozptylu.

Integrace informačních zdrojů se snaží vytvořit mechanismy, kterými by zajistila efektivní využití informací v již zmíněné posloupnosti: nalezení vhodného zdroje pro vyhledávání, vyhledání informace o dokumentu a získání dokumentu. Integrace se obecně nesnaží slučovat všechny informační zdroje do jediného centrálního systému, protože to není většinou výhodné, ani globálně možné (dnešní informační portfolio je díky sítím ohromné a není ve většině případů omezeno na informační zdroje jediného subjektu; různé informační zdroje tak vlastní nebo provozují různé, často navzájem nezávislé subjekty). Integrační řešení pracují s informačními zdroji podle specifických potřeb a jeden zdroj může být integrován do několika různých nezávislých řešení odlišným způsobem. Sloučením zdrojů do centrálního systému by tato možnost zmizela. Při integraci informačních zdrojů je tedy často výhodné respektovat odlišnosti

dílčích prvků, nechat je fragmentované a snažit se hledat a zavádět prostředky pro **vzájemnou komunikaci a výměnu dat** formou standardů a adaptérů. Integrace se tak snaží vytvořit fungující integrované prostředí, ve kterém dochází k „bezešvému“ **propojení** sekundárních informací o dokumentech s vlastními primárními dokumenty, kde je možné komunikovat se všemi zdroji pomocí **jediného dotazovacího jazyka** a kde jsou výsledné záznamy a/nebo dokumenty na výstupu **prezentovány ve sjednocené a standardní formě**. Významným přínosem integrace je také **systematické začlenění** heterogenních zdrojů do jednoho IS, což hraje úlohu nejen evidenční a organizační, ale i monitorovací z hlediska využití zdrojů a akvizice.



Obr. 3 - Funkce ideálního integračního systému z pohledu uživatele

Při tvorbě integračního řešení je třeba mít na paměti, že se jednotlivé složky (dílní informační zdroje) mohou rychle vyvíjet a měnit, je proto třeba budovat ho jako **univerzální a otevřený systém** s jednoduchou architekturou, který bude pružně reagovat na změny dílčích zdrojů, aniž by bylo ovlivněno chování celého systému, zejm. z pohledu koncového uživatele. Tento požadavek nabývá ještě většího významu, když jsme nuceni kromě vlastních **lokálních** informačních zdrojů integrovat i zdroje **externí**, které pracují na informačních systémech, jejichž architekturu nemůžeme ovlivnit.

Integrační řešení využití informačních zdrojů musí podporovat **dosažení strategických cílů** organizace. V případě knihoven, které nejsou přímými konzumenty informací ale jejich zprostředkovateli, je hlavním strategickým cílem uspokojování informačních potřeb uživatelů. Chování veřejných knihoven není motivováno tvorbou zisku jako v komerčních podnicích a zdánlivě mají své roční rozpočty zajištěny. Úspěšnost v uspokojování informačních potřeb uživatelů a udržení jejich přízně však hraje významnou úlohu při zdůvodňování další existence knihoven. Jinými slovy to znamená, že je třeba nabízet a integrovat takové informační zdroje a služby, které jsou pro cílové uživatele relevantní a které jsou poskytovány s přidanou hodnotou. S integrací proto souvisí i další pracovní procesy, v tomto případě zejm. akvizice informačních zdrojů (více informací o akvizici zdrojů - viz kapitola 2.9.3) a navazující služby (více informací o propojení informačních zdrojů a služeb - viz kapitola 2.8).

Integrační řešení by mělo být vyvíjeno na základě **jednotné metodiky, koncepce a soustavy pravidel**. Mělo by maximálně využívat platných standardů, pokud je dílní informační zdroje alespoň částečně podporují. Metodika musí vycházet z procesu projektování integračního IS (více informací o projektování a etapách integrace - viz kapitola 1.6).

Po celou dobu budování a následného provozování systému pro integraci informačních zdrojů je nutné neustále **sledovat kvalitu** všech prvků systému a procesů. Zvláště u větších a metodicky složitějších projektů je nanejvýš vhodné nasadit systém řízení jakosti jako základní nástroj pro

efektivní fungování IS pomocí identifikace a řízení mnoha vzájemně propojených procesů (více informací o řízení kvality při integraci informačních zdrojů - viz kapitola 4).

## 1.4 RIZIKA INTEGRACE

Kromě přínosů integrace informačních zdrojů, z nichž některé jsme zmínili výše, je třeba znát i rizika, která jsou s tvorbou a provozem integračního řešení spojena.<sup>49</sup>

### 1.4.1 Vysoké náklady integrace

Jako u každého projektu tvorby IS je třeba i u budování integračního řešení zvážit, jaké náklady budou s projektem spojeny, zejm. s ohledem na legislativu, koncepci a přípravu konverzních a propojovacích mechanismů. Celý projekt integrace musí být investičně i provozně realizovatelný.

### 1.4.2 Větší složitost integračního systému

Integrační systém je zpravidla složitější než připojené dílčí informační zdroje, přináší sebou proto vyšší nároky na projektování a přípravu řešitelů. Vyšší personální nároky se projeví i při implementaci a provozu systému. Snaha o realizaci příliš rozsáhlých projektů vede často k nezdaru, proto je vhodné rozdělit velký projekt na několik menších projektů s reálně dosažitelnými cíly.

### 1.4.3 Závislost na externích zdrojích

Pokud jsou do integračního systému zapojeny informační zdroje, které jsou provozovány externě jinými subjekty, nemáme nad nimi kontrolu a může docházet nekontrolovaně k problémům se:

- stabilitou (když je zdroj nedostupný z technických důvodů provozovatele)
- kvalitou (nepřesné nebo nestandardní vyhledávání, úroveň zpracování informací)
- seriózností (neaktuálnost a nepravdivost informací)

V případě placených zdrojů, na které se vztahují určité smluvní garance, mohou být rizika ošetřena možností reklamace.

Při vývoji integračního systému je výhodné používat standardizované a obecně rozšířené technologie, které je možné v případě potřeby nahradit. Snížíme tak riziko spojené se závislostí na dodavatelích, kteří mohou z nejrůznějších důvodů přestat spolupracovat (ukončení vývoje technologie, opuštění trhu, krach firmy).

### 1.4.4 Větší a rychlejší dopad případných výpadků integračního systému

Systém pro integraci informačních zdrojů představuje centrální prvek, který v případě výpadku znemožní uživatelům práci se všemi integrovanými informačními zdroji. Výpadky mohou být způsobeny havárií hardware či software, virovým napadením, problémy se síťovým připojením, lidským faktorem apod. Tato rizika můžeme snížit např. sofistikovaným systémem zálohování dat, záložním napájecím zdrojem, duplikací hardware (RAID, záložní síťové připojení), systémy pro síťovou bezpečnost (firewall, proxy, antivirové systémy, ověřovací a šifrovací služby) atd. Chybám způsobené lidským faktorem můžeme předcházet důsledným proškolením pracovníků a zajištěním jejich odbornosti. Je také nezbytné zavést pravidla přihlašovacích práv a uživatelských oprávnění na různých úrovních práce se systémem spolu s nasazením platných provozních předpisů.

### 1.4.5 Hrozba průměrnosti

Různě velký pokrok a úroveň technologií dílčích informačních zdrojů může při integraci do jednotného systému vést k průměrnosti, ve které se ztrácí kvalitní funkcionality proprietárních řešení. Například pokročilý informační zdroj podporující hledání podle 12 vyhledávacích bodů může být po zapojení do integrovaného IS prohledatelný pouze podle 6 vyhledávacích bodů kvůli univerzálnímu použití vyhledávací masky pro všechny připojené zdroje.

### 1.4.6 Nedostatečné zaškolení uživatelů a malá propagace

Integrační IS sice sjednocuje proprietární řešení dílčích informačních zdrojů, ale sám přináší často nově pojaté uživatelské rozhraní, se kterým se musí uživatelé naučit pracovat. Přestože snahou vývojářů je vytvořit intuitivní a uživatelsky přívětivé prostředí, ne každému uživateli bude od počátku srozumitelné. Podceňování zaškolení uživatelů, vytvoření kvalitních nápověd, průvodců a vzdělávacích kurzů, stejně jako nedostatečná propagace nového IS často zbytečně snižují pozitivní dopad implementace a mohou vést až k úplnému nezdaru projektu.

## 1.5 SLOŽKY A ÚROVNĚ INTEGRACE INFORMAČNÍCH ZDROJŮ

Z hlediska knihoven a jim podobných organizací, které zajišťují informační služby svým uživatelům a zprostředkovávají tak přístup k informacím, můžeme řešení integrace informačních zdrojů označit obecně jako **informační bránu**. Informační brána představuje místo, které slouží pro ucelený přístup k dílčím informačním zdrojům. Může být pojata velmi jednoduše i značně komplexně s bohatou škálou služeb, podle úrovně integrace a začleněných složek.

Na základě vlastností informačních zdrojů popsaných v kapitole 1.1 můžeme definovat **prvky**, které mohou tvořit základní nabídku zdrojů a služeb v informační bráně:

- lokální katalogy klasických dokumentů (vlastní klasický fond)
- externí katalogy klasických dokumentů (klasický fond jiných subjektů)
- nejrůznější bibliografické či citační databáze jako pomocné nástroje (např. ČNB)
- referátové oborové a speciální databáze
- nejrůznější faktografické databáze (např. encyklopedie)
- sbírky elektronických dokumentů (např. archivy e-printů či článkové databáze)
- volný prostor internetu
- přidané služby (např. elektronické dodávání dokumentů, uživatelské účty)

Součástí může být i velké množství dalších komponent a služeb, které podrobněji popisuje kapitola o portletech 2.10.2.

### 1.5.1 Přidaná hodnota

Pro informační brány je důležité, jakou **přidanou hodnotu** uživatelům přinášejí. Přidaná hodnota představuje tu kvalitu informační brány, kterou by uživatelé ztratili, kdyby k dílčím zdrojům přistupovali přímo. Vytváří tak výsledný efekt integrace informačních zdrojů z pohledu koncových uživatelů. Podle míry přidané hodnoty při integraci informačních zdrojů můžeme informační brány rozdělit na tyto typy:

- seznamy externích odkazů - rozcestníky (odkazují na jiné IS, lokální i externí)
- vyhledávání v lokálních a/nebo externích zdrojích z jednoho místa
- paralelní vyhledávání v lokálních a/nebo externích zdrojích
- paralelní vyhledávání doplněné o navazující informační služby

**Seznamy externích odkazů** jsou většinou řazeny do klasifikačních, zejm. předmětových kategorií a zajišťují tak uživatelům orientaci a přehlednost. Často jsou také podle předem určených hledisek selektovány, zejm. podle kvality obsahu, a uživatelé tak mají zajištěnu potřebnou informační úroveň. Tvorbou seznamů odkazů může vycházet nejen z principu vytipování zdrojů a akvizice, ale i z citačního principu. Odkazy mohou být doplněny o krátké anotace nebo charakteristiky. Významným rysem seznamů odkazů je také pravidelné ověřování dostupnosti zdrojů (např. ověřování platnosti odkazů URL). Aby byly seznamy stále aktuální a relevantní, je třeba systematicky a pravidelně vyhledávat nové informační zdroje a odstraňovat neplatné či nerelevantní. Přidaná hodnota této služby tedy spočívá v zajištění vyšší úrovně kvality a relevance nabídky, zajištění platnosti a dostupnosti zdrojů a v třídění do přehledného systému klasifikace.

**Vyhledávání v informačních zdrojích z jednoho místa** je služba, která umožňuje z jednoho rozhraní vyhledávat informace ve více informačních zdrojích, na rozdíl od paralelního vyhledávání ale pouze jednotlivě. Hlavním přínosem této služby je, že uživatel přistupuje k připravené nabídce zdrojů, které jsou často předmětově tříděny, a může v libovolném z nich vyhledávat v jednotném uživatelském prostředí. Výsledky mohou být navíc převáděny do

jednotné podoby. Přidaná hodnota se tedy oproti seznamu externích odkazů ještě zvyšuje o sjednocení dotazovacího jazyka a odbourání přechodů do rozhraní jiných IS.

**Paralelní vyhledávání ve zdrojích** navazuje na předchozí službu a umožňuje vyhledávání ve více informačních zdrojích současně. Uživatel tak nemusí postupně posílat dotazy do jednotlivých zdrojů, ale vyhledávání provede najednou v jeden okamžik. Služba dokáže zpracovat jednotlivé výsledky a často nabízí navazující funkce jako je deduplikace a slučování. Přidaná hodnota se touto službou navyšuje o úplné sjednocení informačních zdrojů na vstupu i na výstupu vyhledávání, takže uživatel pracuje s více zdroji jako s jedním.

Paralelní vyhledávání může být doplněno o **navazující služby**, které uživateli rozšiřují možnosti, jak získat primární dokumenty či digitální objekty nebo související informace (elektronické dodávání dokumentů, MVS, rezervace, ověření dostupnosti, odkazy na citované nebo citující dokumenty apod.). Významnou a velice informačně cennou službou může být také budování hypertextových odkazů mezi dokumenty a záznamy na bázi bibliografických údajů a citačních vazeb. Mohou to být také služby umožňující jiné formy vyhledávání nebo různé transformace informací podle potřeb uživatele (vazby na uživatelskou schránku, vytváření osobních citačních a bibliografických seznamů, automatické alerty a SDI, vazby na RSS, vyhledávací profily apod.). Škála služeb může být velice široká, a proto i další navyšování přidané hodnoty je touto cestou v podstatě neomezené.

Přidanou hodnotu ale nezvyšuje pouze míra vlastní integrace zdrojů, ale také samotné pracovní prostředí. Uživatel využívající informační bránu může například získat:<sup>42</sup>

- vyšší rychlost při výběru zdroje i při vlastním vyhledávání
- rovnocenný přístup ke zdrojům a službám bez ohledu na věk, handicap a jiné faktory
- využitelnost na dálku mimo prostory knihovny
- anonymní přístup k informačním zdrojům ale i možnost přístupu přes ID
- předdefinované dotazy, předvyplněné formuláře a uživatelské nastavení
- časově neomezená dostupnost služeb bez ohledu na otvírací hodiny provozovatelů
- samoobslužný princip ale i možnost asistence (e-mail, telefonní podpora)
- vícejazyčné prostředí
- podpora přístupu přes různé technologie, např. PC, PDA, WAP atd.
- bezpečnost a soukromí při práci, ochrana osobních informací podle legislativy
- možnost zpětné vazby zbavené komunikačních bariér při osobním styku

### 1.5.2 Zaměření na uživatelské skupiny

Při projektování systému pro integraci informačních zdrojů je třeba jako jednu z prvních věcí vyřešit otázku, jakým skupinám uživatelů bude informační brána sloužit. Je tím významně ovlivněn záběr a šíře nabídky informačních zdrojů a služeb, které by měla brána obsahovat.

Organizace může budovat informační bránu pouze pro účely vlastních zaměstnanců nebo členů, případně svých zákazníků, pak mluvíme o **podnikové informační bráně**. Příkladem může být informační systém knihovny rozdělený na služební část určenou pro knihovníky/informační profesionály a na „klientskou“ část pro čtenáře. Obsah brány je dán výhradně potřebami dané organizace.

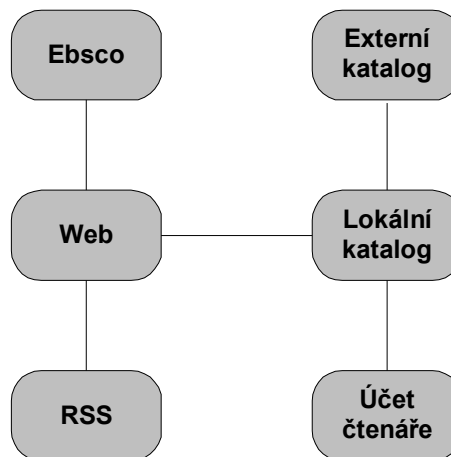
Informační brána nemusí vznikat pouze pro potřeby jednoho subjektu, ale může být zaměřena na určitý předmět zájmu, např. vědní obor nebo odvětví. V takovém případě mluvíme o **oborové informační bráně** a jejími uživateli jsou potencionálně všichni lidé, kteří se profesionálně, studijně nebo laicky zajímají o daný obor. Významným znakem takové skupiny uživatelů je vyšší míra anonymity, pokud brána nepožaduje povinně autentikaci.

Nejširšímu okruhu uživatelů slouží **univerzální informační brána**, která se snaží integrovat informační zdroje ze všech oborů lidské činnosti. Uživatelé téměř vždy přistupují k bráně anonymně, i když může brána poskytovat i personalizované služby. Univerzální informační brány musí sloužit uživatelům s různými informačními potřebami.

Odlišný typ představuje **systémová informační brána**, která neslouží potřebám uživatelů ale jiných IS. Jejím úkolem je obecně propojit a umožnit výměnu dat a služeb mezi dvěma a více IS. Typické je, že není vybavena uživatelským rozhraním. Příkladem může být systém pro integraci informačních zdrojů pro účely stahování bibliografických záznamů přes protokol Z39.50.

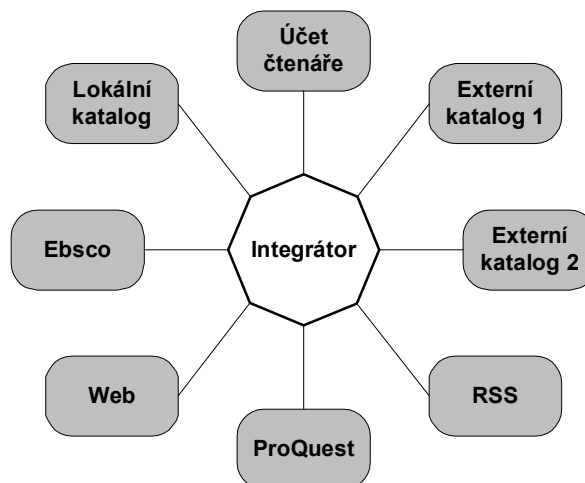
### 1.5.3 Modely propojování informačních zdrojů

**Propojování point-to-point** představuje způsob integrace, kdy jsou jednotlivé prvky systému (tedy každý informační zdroj nebo služba) propojeny na principu vazby jeden s jedním. Vznikají proprietární adaptéry pro každé spojení, takže lze využívat data i z jiných zdrojů, ale nelze k nim přistupovat z jediného místa. Často také nejsou vytvořeny vazby mezi všemi prvky. Tento způsob propojování je vhodné použít v prostředí, kde je relativně nízký počet informačních zdrojů a kde je snadné vytvářet proprietární adaptéry díky technickým vlastnostem dílčích informačních zdrojů. Typické point-to-point prostředí ukazuje následující náčrt.



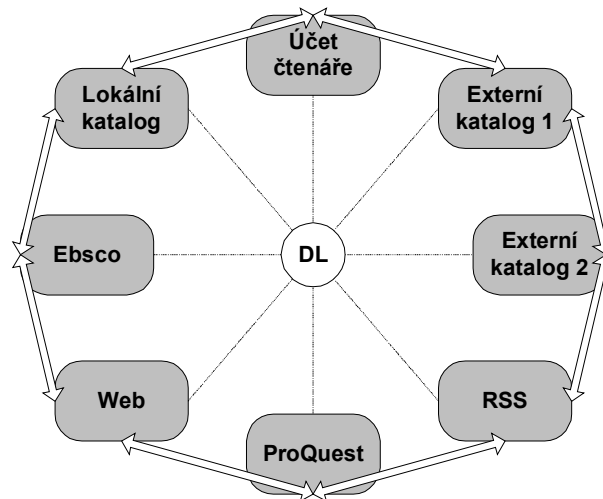
Obr. 4 – Proprietární vazby dílčích zdrojů

**Propojování pomocí centrálního prvku** předpokládá existenci sdíleného místa, odkud jsou posílány dotazy, kam jsou posílána data (např. výsledky dotazů) z jednotlivých zdrojových aplikací a kde dochází k jejich transformaci. Zcela jsou odstraněny vazby mezi jednotlivými informačními zdroji a potřeba vytvoření rozhraní se redukuje pouze na vazbu mezi centrálním prvkem a příslušným zdrojem. V případě standardizovaného integračního nástroje se integrace dále výrazně zjednoduší. Pokud dojde k náhradě některé části systému jinou aplikací, je nutné řešit vždy pouze rozhraní mezi integračním nástrojem a novým zdrojem. Všechny zbývající vazby zůstávají nedotčeny. V rámci komunikace jsou využívány takové protokoly a služby, které vyhovují jednotlivým informačním zdrojům. Tento způsob propojování je velice flexibilní, protože jednotlivé informační zdroje se navzájem neovlivňují. Je však poměrně náročný na vlastnosti integračního prvku.



Obr. 5 - Přístup z jediného místa ke všem zdrojům

**Propojování dynamickým linkováním** využívá inteligentní a kontextově orientovaný prvek, který obsahuje informace o vyhledávacích a komunikačních schopnostech všech dílčích informačních zdrojů a služeb. Při práci s jedním zdrojem dokáže na základě výsledků nabídnout vazby na další zdroje podle dané situace. Např. při práci v lokálním katalogu uživatel získá odkaz na žádanku o meziknihovní výpůjčku, možnost využít externí služby elektronického dodávání dokumentů pro vyhledaný dokument nebo zobrazit citace daného díla v citačním rejstříku.



Obr. 6 - Dynamický linker (DL) pro generování vazeb

#### 1.5.4 Úrovně integrace

Integrace informačních zdrojů je složitá v tom, že může být nutné integrovat jak **námi provozované**, většinou lokální zdroje, tak i informační zdroje **provozované jinými subjekty**. Zatímco u našich zdrojů můžeme více či méně definovat architekturu a chování, u zdrojů, které provozuje jiný subjekt, často nemáme možnost vlastnosti nijak ovlivnit. Míra vlivu je potom dána formou spolupráce nebo obchodních dohod. Obecně platí, že možnosti integrace vlastních zdrojů jsou daleko větší, než je tomu u zdrojů provozovaných jinými subjekty. Typickými příklady vlastních zdrojů v knihovnách jsou katalogy, vytvářené bibliografické a faktografické databáze, digitální sbírky rukopisů a starých tisků apod. Příklady zdrojů provozovaných jinými subjekty jsou všechny komerčně provozované zdroje, jako je Ebsco, ProQuest nebo Web of Science, katalogy jiných knihoven apod. Někde mezi tím stojí zdroje budované spoluprací více subjektů, kde mohou jednotlivé subjekty významně ovlivnit řešení, ale kde musí také často přistoupit ke kompromisům.

Při integraci informačních zdrojů můžeme definovat typy integrace na následujících úrovních:<sup>50</sup>

- datová integrace
- integrace aplikací
- integrace procesů
- integrace uživatelských rozhraní
- integrace metodická
- integrace technologická

**Datová integrace** představuje řešení, kdy několik aplikací využívá společné datové úložiště, čímž se zamezuje duplikovanému ukládání a zpracování dat. Příkladem může být projekt národních autorit NK ČR, kde mnoho českých knihoven zapisuje autoritní záznamy do jedné centrální databáze, odkud je poté pomocí vazeb propojují s dalšími údaji ve svých aplikacích. Významným efektem datové integrace je jednotnost údajů daná jednotnou strukturou databáze a neduplicitní prostředí.

**Integrace aplikací** je založena na propojování jednotlivých funkcí a sdílení dat různých aplikací. Aplikace musí být schopny volat a spouštět funkce externě. Na tomto principu je



založena řada technologií, jako např. webové služby používající k volání funkcí hodnoty kódované do URL nebo dotazy XML.

**Integrace procesů** navazuje na integraci aplikací, integruje však kompletní sady funkcí zajišťující celé procesy zpracování a zpřístupňování informačních zdrojů. Integrace procesů předpokládá, že je IS pevně spjat s procesy organizace a že organizace dosáhla vysokého stupně automatizace. Příkladem integrace procesů může být projekt centrální katalogizace, ve kterém se tři velké knihovny (NK ČR, MZK a VKOL) dohodly, že sjednotí své procesy katalogizace a budou vytvářet společný centrální katalog společnými postupy.

**Integrace uživatelských rozhraní** sjednocuje vzhled a ovládání rozhraní jednotlivých aplikací. V dnešní době došlo k velkému pokroku v integraci díky grafickým prvkům OS, zejm. MS Windows, a díky masivnímu využívání webového prohlížeče jako uživatelského klienta. Přesto se uživatelská rozhraní mezi sebou odlišují vlivem rozdílných funkcí, struktury a designu. Úplná integrace uživatelských rozhraní spočívá v úplné eliminaci dílčích rozhraní a vytvoření nového, společného uživatelského rozhraní v rámci integračního řešení. Příkladem může být projekt Jednotné informační brány.

**Integrace metodická** představuje sjednocení pracovních postupů a metodik zpracování a zpřístupňování informačních zdrojů. Tato úroveň je v oblasti knihovnictví důvěrně známá, protože metodické postupy jsou dobře koordinovány a sjednocovány. Příkladem úspěchů tohoto typu integrace je bezesporu celá sada obecně přijímaných předpisů pro zpracování dokumentů (MARC, AACR2, ISBD).

**Integrace technologická** představuje návrat k centralizovaným řešením zpracování dat. V globálním měřítku, kde může být IS distribuován prakticky po celém světě v podobě samostatných heterogenních prvků, je třeba řešit integraci informačních zdrojů technologicky pomocí mezinárodních standardů a vzájemných dohod (znakové sady, datové profily, formáty).

### 1.5.5 Typické skupiny informačních zdrojů z hlediska integrace v knihovnách

Knihovny mají přístup k většímu počtu informačních zdrojů než samotný uživatel; to knihovnám přináší významnou výhodu a je předpokladem k poskytování těžko nahraditelných a cenných služeb.

První velkou oblastí informačních zdrojů je **fyzický knihovní fond**, který knihovna vlastní. Jedná se o klasické i elektronické dokumenty fixované na fyzických nosičích, které nejsou digitálně přístupné online, např. knihy, časopisy, normy, LP, CD, CD-ROM, noty apod. Protože se jedná o fyzicky uložený fond, nelze ho integrovat přímo, ale prostřednictvím elektronické evidence (OPAC) a pomocí zprostředkovatelských služeb, např. elektronického dodávání dokumentů. Vzácné části fondu se postupně digitalizují, zejm. historické a vzácné sbírky, a zpřístupňují online.

Druhou část tvoří vlastní nebo zakoupené digitální sbírky a databáze, které tvoří **elektronický knihovní fond** (často označovaný jako digitální knihovna) dostupný uživatelům pouze online. Řada knihoven vytváří cenné faktografické databáze nebo sbírky zdigitalizovaných starých tisků, rukopisů či historických časopisů. Kupují také licence plnotextových nebo bibliografických databází komerčních vydavatelů a elektronické knihy (eBooky). Pro tyto informační zdroje je typické, že informace v nich obsažené jsou alespoň v minimální míře organizovány a popsány.

Třetí, téměř bezbřehou oblast informačních zdrojů představuje tzv. **mělký web**, tedy volný prostor internetu tvořený statickými dokumenty a statickými sídly, které mají pevnou adresu URL. Mohou to být dokumenty a soubory nejrůznějších formátů. Tato část internetových zdrojů je relativně dobře zmapovatelná běžnými indexovacími a vyhledávacími nástroji, jejichž technologie jsou založeny na indexování pomocí sledování odkazů URL. Informační zdroje spadající do této skupiny jsou díky pevné adrese URL „uchopitelné“.

Čtvrtou oblast, která je z hlediska využití a integrace nejproblematictější ale informačně velice cenná, tvoří dynamicky generované dokumenty a prohlídatelné databáze na internetu. Označuje se často jako neviditelný nebo **hluboký web**. Informační zdroje této skupiny nelze většinou automaticky indexovat a prohlédávat, takže se uživatelé nemohou o obsahu těchto zdrojů dovědět pomocí běžných vyhledávačů. Pro přístup k dynamickému obsahu je třeba zadávat

vstupní data nebo použít vyhledávací jazyk rozhraní. Informační zdroje databázového typu také často vyžadují autentikaci, kdy je pro vstup do systému nutné zadat ID a heslo.

Knihovny mohou integrovat všechny čtyři skupiny zdrojů dohromady a využívat je jako jeden informačně silný zdroj v podobě knihovnických služeb. Základem je vlastní knihovní fond obohacený o fondy dalších knihoven na základě spolupráce MVS. Tyto fondy mohou být nabízeny na dálku v podobě elektronického dodávání dokumentů nebo lokálně. Mělký web, ke kterému mají přístup všichni uživatelé nezávisle na knihovně, mohou knihovny obohacovat o přidanou hodnotu v podobě klasifikace, předselekcce a organizace. Oblast databází je doménou knihoven z hlediska nákupu licencí či akvizice a i zde mohou přidávat hodnotu díky popisu a klasifikaci. Zdroje mohou sjednocovat z hlediska rozhraní, dotazovacích jazyků i výstupů a umožňovat jejich paralelní prohledávání. Na všechny informační zdroje mohou také navazovat další odborné informační služby, zejm. rešeršní, analytické a doručovací.

## 1.6 PROJEKTOVÁNÍ IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ

### 1.6.1 Metodologie projektování IS

Při projektování IS obecně, a tedy i IS pro integraci informačních zdrojů, je nutné zavádět normy a standardizovat postupy, protože se jedná o velmi složitý proces, který probíhá relativně dlouhou dobu a podílí se na něm mnoho lidí z různých oblastí pracovní činnosti. Integrační systém navíc spojuje často velmi heterogenní dílčí systémy, takže standardizace je základním předpokladem efektivního spolupráce s okolím. Standardizace projektování IS se odráží v postupech dané organizace nebo se odvíjí od požadavků subjektu, který poskytuje finanční prostředky na realizaci projektu (např. grantové agentury).

Teorie a metodologie projektování informačních systémů jsou vždy značně ovlivněny **technologiami**, které jsou v dané době k budování IS k dispozici. Jsou ovlivněny i aspektem **způsobu řízení dané organizace**, která IS buduje, protože projektování s řízením úzce souvisí. Řízení organizace může projektování IS značně determinovat. Nejedná se přitom pouze o vliv „shora“ od vedoucích pracovníků a managementu, ale i „zezdola“ od budoucích uživatelů IS. Pokud vedení organizace nepodporuje spolupráci na projektování IS se zástupci budoucích koncových uživatelů, může to IS negativně ovlivnit.

Pojetí projektování IS z pohledu řízení bývá nejčastěji ovlivněno aspektem čistě **ekonomickým** (na IS se pohlíží jako na nástroj k zajištění/zvýšení efektivity a produktivity organizace čistě z hlediska finančního přínosu/úspor) a aspektem **kvality** (u IS se zkoumá, jakým způsobem bude zajištěna/zvýšena kvalita činností a procesů v organizaci). Oba aspekty se kombinují, ale často jeden z nich převažuje podle způsobu řízení organizace. Bohužel právě finance často determinují způsob, jakým se bude IS budovat.

U projektů integračních IS je důležité, v jakém **stádiu automatizace** se nacházejí dílčí informační zdroje, se kterými bude projektovaný IS spolupracovat. Stádium automatizace se v připravovaných projektech IS významně odráží, protože definují prostředí a výchozí situaci. Nejběžněji se pro vývojová stadia automatizace podniku používá klasický Nolamův model, který definuje stav jako<sup>29</sup>:

1. počátek (málo výpočetní techniky, minimální plánování)
2. rozšiřování (technologie se rozšiřují, plánování stále nedostatečné)
3. řízení (důraz na plánování a řízení automatizace)
4. integrace (začátek řízení datových zdrojů)
5. správa dat (řízení datových zdrojů, ale stále volnost ve vývoji systémů)
6. zralost (aplikace zcela zaplňují všechny procesy, nové aplikace pouze ty s velkým přínosem)

Je také důležité, jestli projekt řeší budování zcela nového IS, nebo má za cíl transformovat stávající IS či ho inovovat. Také přístup řešitelů k řešení může být konvenční nebo novátorský - nelze říci, že konvenční řešení je špatné; v řadě případů konvenční řešení postačuje, protože plně vyhovuje potřebám.

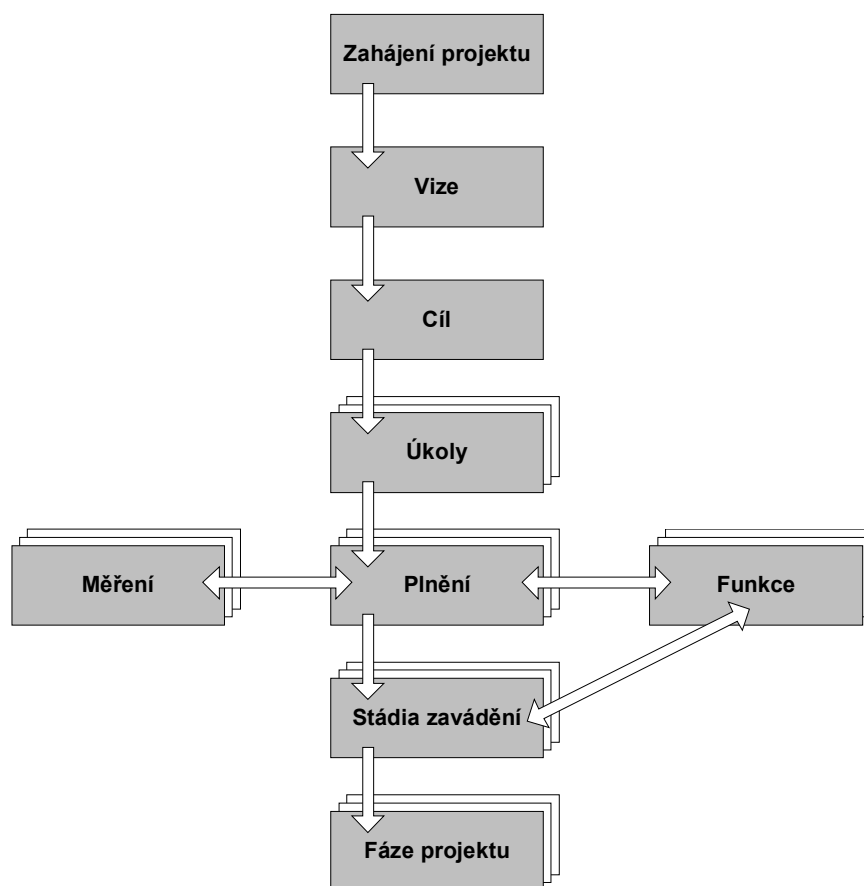
Na stav automatizace můžeme také pohlížet z hlediska **spolupráce a integrace**. K tomu můžeme využít modelu Huff-Mumro-Martinové<sup>29</sup>:

1. izolace (žádná výměna dat mezi aplikacemi)
2. manuální integrace (výměna dat ale pouze ručně na úrovni souborů)
3. automatická integrace (sdílená úložiště dat, replikace, synchronizace)
4. distribuovaná integrace (přístup k úložištím dat bez ohledu na místo uložení)

Z výše uvedených aspektů vyplývají obecné zásady, které by se měly v metodách projektové práce odrážet. V první řadě je třeba zajistit **podporu ze strany vedení** a zajistit **zapojení koncových uživatelů do návrhu**. Vedení by nemělo při projektování vystupovat jako oponent ale jako spoluředitel. Projekty IS by se měly **orientovat na primární cíle** a hlavní problémy organizace - IS má být nástrojem pro dosahování cílů spolu s dalšími nástroji, kterými organizace disponuje. Výstupem každé etapy projektování by měly být **dokumenty** (mezi které patří i schéma architektury IS), které by měly být předmětem schvalování ze strany jmenované komise a vedení organizace. Pomocí těchto dokumentů je možné včas modifikovat a řídit projekt a předcházet tak pozdějším velkým a nákladným změnám v IS. S tím souvisí i potřeba **průběžného ověřování a testování** návrhu během celého procesu projektování. Ověřováním se kontroluje harmonogram, dosahování cílů a zajištění kvality. Analýza a návrh se v jednotlivých etapách **postupně prohlubují**, čímž se zabrání příliš podrobnému návrhu na začátku projektu, který by se později měnil, a naopak je zajištěna možnost průběžné modifikace v průběhu. Každá etapa by měla **sledovat všechny složky a pohledy** na IS, aby byl návrh úplný. Jedná se zejména o analýzu dat (typy dat a jejich struktura), funkcí a procesů, o analýzu organizační struktury organizace a potřeb ji změnit, o sociální, psychologická a ekonomická hlediska a o analýzu technologií. V neposlední řadě je třeba respektovat **otevřenost metodologie** projektování dalšímu vývoji a možnost jejich úprav podle stávajících a měnících se situací v organizaci a v jejím okolí.

Velice důležité je stanovit **vizi**, která slouží jako hlavní motivační síla celého projektu. Vize ztělesňuje základní filozofii a poslání organizace. Z vize by měl vycházet konkrétní **cíl projektu** odrážející předmět a určení integrace: **co** chce organizace integrovat a pro **koho**. Orientace na koncového uživatele by měla být samozřejmostí (nejedná-li se o projekty integrace na systémové úrovni IS). Typy koncových uživatelů, definované podle věku, národnosti, lokace, zkušenosti atd., v důsledku významně ovlivňují technické řešení, např. aspekt vícejazyčnosti, úroveň rozhraní podle stupně náročnosti apod. Dobrou pomůckou při stanovování cíle projektu je příprava **možných scénářů** chování koncových uživatelů při práci s informačními zdroji. Rozborem scénářů lze získat řadu cenných poznatků, zejm. popis různých informačních situací, požadavky na personalizaci a přednastavení různých částí IS, požadavky na spolupráci různých modulů IS mezi sebou atd. Z cíle projektu se na základě analýzy stanoví jednotlivé **úkoly**. U každého požadavku je vhodné specifikovat stav, kdy se úkol považuje za splněný, a termín, do kdy je nutné úkol splnit. Termíny vytvářejí časový harmonogram plnění projektu. Z jednotlivých splněných úkolů vycházejí návrhy a funkce, které jsou postupně ověřovány a upravovány až do stavu, který plně vyhovuje potřebám projektu. S laděním je spojena řada kontrolních **měření** a výsledkem pak jednotlivé **fáze implementačních stádií**. Celý postup projektování může být ovlivněn nasazeným typem systému řízení jakosti (více informací o řízení jakosti, viz kapitola 4).

Při integraci informačních zdrojů, které jsou často externí, je situace komplikovanější, protože je třeba počítat s prvky systému, jejichž chování nemůžeme ovlivnit. Důraz na otevřenost je proto o to větší. Použitá metodologie vychází ze specifických potřeb daného projektu, zvyklostí organizace, popř. ze specifik použitého CASE nástroje.



Obr. 7 – Jedno z možných schémat metodologie projektování

### 1.6.2 Projektové vrstvy

Vývoj, implementace a provoz IS pro integraci informačních zdrojů je proces, který můžeme rozdělit na linii technickou a na linii organizační.

**Organizační vrstva** zajišťuje řídicí rámec projektu, financování a budování fungujícího obchodního modelu, tvorbu projektové strategie a budování vztahů s partnery a dodavateli včetně organizačního zajištění integračních vazeb na úrovni připojovaných informačních zdrojů a služeb. Řada integrovaných prvků může být externí, a zatímco systém je na nich závislý, externí prvky na systému závislé nejsou. Tento fakt významně ovlivňuje proces projektování tohoto typu IS. Integrační vazby nestačí zajistit pouze technicky, ale také organizačně. Organizační vrstva dále zajišťuje kontrolu kvality všech fází projektu i sledování kvality samotných informačních zdrojů a služeb, sleduje soulad s legislativou, stará se o efektivní marketing a propagaci projektu. Vytváří také vize pro pokračování projektu nebo dalších navazujících projektů.

Obchodní model je základním předpokladem trvalého a úspěšného fungování integračního systému. Zaručuje také do jisté míry ochranu vložených investic. Nestačí pouze zajistit finance na vývoj a implementaci IS, ale také na jeho provoz a rozvoj. Z dlouhodobého pohledu mohou být finanční prostředky potřebné na provoz vyšší než náklady na realizaci IS. Řada projektů brzy po realizaci mizí ze světa právě proto, že nejsou zajištěny dostatečné prostředky do dalších let. V prostředí knihoven je problém zajištění obchodního modelu složitý, protože knihovny nejsou komerčně založeny a nejsou také příliš sponzorsky atraktivní. Běžný roční rozpočet knihoven nedovoluje vznik nákladných informačních projektů a většina investic tak většinou plyne z vypisovaných grantů ministerstev nebo EU. Grantové programy, přestože mohou být víceleté, však mnohdy neumožňují financovat provoz - buď vůbec nebo velice omezeně. Existuje však mnoho možností, jak efektivní obchodní model projektu vybudovat. Několik řešení popisuje kapitola 6.1.5.

**Technická vrstva** projektu navazuje na projektovou strategii a reaguje na definované cíle a úkoly. Zajišťuje vlastní projektování IS, jeho vývoj a implementaci. Zabývá se zejm. oblastí dat, funkcí a procesů, tvorbou a výběrem software a volbou hardware. V průběhu projektování vzniká model řešení, který je v jednotlivých fázích analyzován a měřen a který nakonec dospívá do fáze prototypu. Na výstupu každé fáze vznikají dokumenty, které slouží jako základní vazba s organizační vrstvou. Otestovaný model nebo jeho části jsou v konečné fázi implementovány a zpřístupněny uživatelům. Technická vrstva také zabezpečuje zpětnou vazbu od uživatelů a vlastního IS formou statistik, logů a reportů nebo dalších nástrojů.

## 1.7 ETAPY PROJEKTOVÁNÍ A INTEGRACE

IS pro integraci informačních zdrojů můžeme rozdělit na dvě logické části: na **vlastní integrační IS**, který tvoří aplikační základ, a na **jeho obsah**, který tvoří připojené informační zdroje a služby. Základním předpokladem úspěšnosti integračního řešení, který byl již v předchozích kapitolách několikrát zdůrazněn, je maximální otevřenost architektury aplikační části a úplná nezávislost na konkrétních připojených zdrojích.

### 1.7.1 Etapy projektování a implementace integračního systému

Implementace IS probíhá postupně v závislosti na plnění dílčích úkolů projektu. Úkoly jsou seskupeny do na sebe navazujících implementačních stádií a v této řadě je strategicky důležité zvolit okamžik spuštění v ostrém provozu. Pokud tento okamžik nastane příliš brzy, je zahájení provozu provázeno velkou chybovostí a nestabilitou. To může mít negativní vliv na reakci uživatelů a na jejich důvěru. Naopak snaha o nasazení IS až v okamžiku „úplné dokonalosti“ je také strategicky chybná, protože integrační systém je neustále se vyvíjející organismus, který nikdy nedospěje do definitivního, cílovém stavu. Vývoj integračního IS končí až v okamžiku jeho nahrazení jiným IS. Řada projektovaných IS nebyla nikdy nasazena právě proto, že tvůrci nedokázali nasazovat systém po etapách a snažili se o jeho „úplné dokončení“.

Integrační IS je proto třeba nasazovat postupně, vždy když je dokončena a otestována určitá ucelená část funkcionality, která může sloužit uživatelům. Nasazené části systému jsou poté ve fázi provozu, zatímco implementace dalších částí se připravuje. Uživatelé tak mají poměrně brzy k dispozici řešení, které se postupně zdokonaluje. Jednotlivé implementační fáze je vhodné pro uživatele označovat (např. verzováním), aby nebyli zmateni nečekanými změnami v systému. Nasazování verzí (přidávání a změny částí/modulů) se musí realizovat s ohledem na psychologii uživatelů – s příliš častými změnami se nedokáží uživatelé dostatečně vyrovnat a naopak příliš dlouhé prodlevy mezi verzemi zase vzbuzují mezi uživateli pocit, že se projekt nevyvíjí. Celý proces zavádění a uvolňování verzí je třeba podpořit dostatečnou marketingovou oporou.

### 1.7.2 Etapy integrace zdrojů do systému

Jak jsme již zmínili v předchozích kapitolách, informační zdroje se neustále mění (vznikají, zanikají, transformují se), a proto proces jejich integrace nekončí po celou dobu životního cyklu projektovaného IS. Při projektování IS pro integraci informačních zdrojů je nutné vytvořit harmonogram pro postupné začleňování informačních zdrojů a navazujících služeb do systému. Ve fázi provozu je potom nutné zajistit nepřetržitou kontrolu funkčnosti připojených zdrojů, průběžnou akvizici a správu nabídky. Předpokladem neustálé obměny zdrojů je již několikrát zdůrazněná otevřenost a flexibilita celého řešení.

Základní kroky začleňování informačního zdroje do integračního systému mohou být např.:

- akvizice zdroje (solistikovaný výběr s ohledem na informační potřeby uživatelů)
- popis zdroje a jeho klasifikace (zařazení zdroje do přehledného systému třídění)
- adaptace (technické zapojení pro účely vyhledávání a komunikace)
- ladění pro účely paralelního vyhledávání
- napojení na služby dodávání primárního dokumentu
- napojení na další služby a procesy

Tyto kroky představují možné základní administrativní aktivity v systému. Jednotlivé kroky by měly být v systému podpořeny příslušnými funkčními moduly, aby byly automatizované. Výsledkem procesů akvizice, popisu, adaptace a dalších nastavení by mělo být maximální

sjednocení chování dílčích informačních zdrojů na straně uživatelského rozhraní. Jednotlivé aspekty jsou z technického hlediska popsány v kapitole 2 pojednávající o architektuře IS.

## **1.8 SHRNU TÍ**

**Integrace informačních zdrojů se snaží řešit vážné problémy spojené s využitelností bohatého ale heterogenního informačního portfolia, které mají knihovny jako provozovatelé informačních služeb k dispozici. Cílem projektovaných integračních řešení je nabídnout uživatelům prostřednictvím služeb co nejvíce informačních zdrojů, které prošly pečlivou selekcí z hlediska kvality, nabízet je jednotně a organizovaně v přívětivém prostředí, sjednocovat jejich vstupy i výstupy a nabízet na ně navazující služby.**

## 2 ARCHITEKTURA IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ

Integrační systém představuje komplexní řešení propojení heterogenních informačních zdrojů a služeb. Musí nabízet jednoduchá a snadno použitelná rozhraní pro uživatele, ale přitom reagovat na často velmi složité a mnohaúrovňové vztahy dílčích integrovaných prvků. Při projektování integračního systému je proto nesmírně důležité **zvolit vhodnou architekturu**, která umožní a zajistí pokrytí všech požadavků na funkcionalitu, snadnou a rychlou implementaci systému, otevřenost k nově připojovaným zdrojům, flexibilitu k novým a pozměněným požadavkům, stabilitu, bezpečnost, spolehlivost a jednoduchost směrem k uživatelům.

Každý integrační projekt má svá specifika, proto v této kapitole podrobněji prozkoumáme pouze ty části IS, kterých se integrace nejpravděpodobněji dotkne. Po shrnutí základních principů architektury IS pro integraci informačních zdrojů si ukážeme, jak lze integrovat **uživatelská rozhraní** (kapitola 2.2), **autentikační a autorizační mechanismy** s ohledem na **bezpečnost** (kapitoly 2.3 a 2.4), **akviziční procesy** při získávání informačních zdrojů (kapitola 2.5), jak lze **popisovat a organizovat** integrované informační zdroje (kapitola 2.6), jak integrovat **vyhledávání** (kapitola 2.7) a jak lze vyhledávací procesy **propojit s navazujícími službami** (kapitola 2.8). V závěru kapitoly si ukážeme, jak důležité je v integrovaných systémech využívat **zpětnou vazbu** (kapitola 2.9), a jakým způsobem můžeme všechny integrační prvky propojit v **portálu** (kapitola 2.10).

Během životního cyklu integračního systému přicházejí na trh stále nové informační technologie, které pozvolna nahrazují ty starší. V každém okamžiku je na informačním trhu podporováno a nabízeno několik generací mnoha různých technologických řešení pro zpřístupňování informačních zdrojů a jejich integraci. Na měnící se nabídku průběžně reagují provozovatelé informačních zdrojů a služeb, kteří inovují svá technická řešení, čímž dochází k dynamickým změnám prostředí, ve kterém integrační systém pracuje. U IS pro integraci informačních zdrojů je proto třeba maximálně zdůraznit potřebu **otevřenosti** řešení, aby bylo možné průběžně integrovat nové informační zdroje a služby, a potřebu **flexibility**, aby bylo možné pozdější změny provádět za provozu bez omezení práce uživatelů.

V tomto oddíle si ukážeme, jaké aspekty nejvýrazněji ovlivňují architekturu IS pro integraci informačních zdrojů a jaké problémy je třeba při návrhu nejčastěji řešit.

### 2.1 ZÁKLADNÍ PRINCIPY ARCHITEKTURY IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ

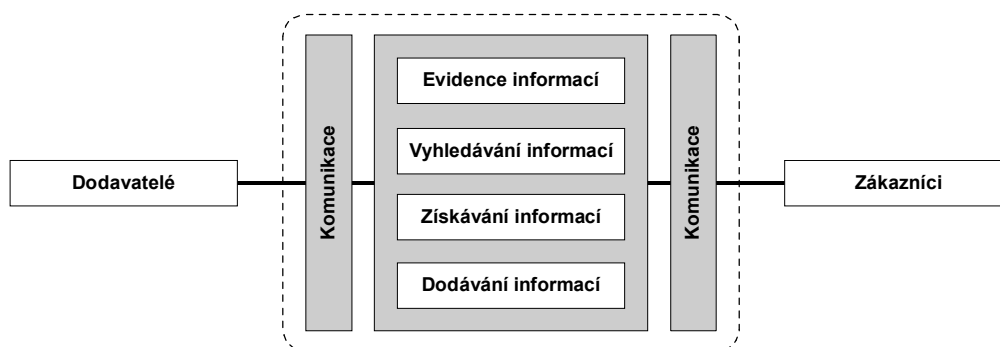
Architektura IS vytváří základní rámec řešení IS s jasně a přehledně definovanými stavebními prvky a vztahy mezi nimi. Z navržené architektury vycházejí konkrétní hardwarová, softwarová a organizační řešení, která naplňují funkční, procesní a datové vlastnosti jednotlivých stavebních prvků. Architektonický model IS slouží v jednotlivých projektových fázích k formulaci představ o IS, k analýze a návrhu řešení, k definování dílčích projektů a úkolů a v pozdějších fázích také k modelování možných změn a plánování dalšího vývoje IS.

Tvorba architektury IS je metodicky spjata s celkovou metodikou projektování IS. Takových metodik existuje celá řada a jejich výběr závisí na zvyklostech organizace, pokynech nadřízeného subjektu či strategického partnera, na povaze a složitosti projektu atd. Obecné metodiky lze také modifikovat podle interních potřeb organizace nebo dokonce konkrétního projektu. V ČR se v posledních letech významně prosadila moderní metodika **MDIS** (Multidimensional Development of Information System) vyvinutá na VŠE v Praze<sup>49</sup>, která byla úspěšně použita v řadě nedávno realizovaných projektech ve státní i podnikové sféře. Jednou z hlavních výhod MDIS je skutečnost, že není nijak fixována na softwarová a hardwarová řešení, ani na konkrétní modelovací nástroje typu CASE. Možné typy architektur IS proto v této práci ukážeme na rozlišovacích modelech metodiky MDIS. Přesná podoba architektury integračního systému samozřejmě vychází z konkrétních projektů a nelze ji obecně popisovat. Příklad možného řešení ukazuje kapitola 5.

### 2.1.1 Globální architektura IS pro integraci informačních zdrojů

Globální architektonický model IS pro integraci informačních zdrojů představuje celkové schéma IS s vyznačenými dílčími komponenty a jejich vzájemnými vazbami, který zajistí uspokojování informačních potřeb uživatelů. Slouží pro účely analýzy a znázorňuje cílový stav daného řešení. Na základě globální architektury pak definujeme jednotlivé dílčí architektonické modely jako podrobnější návrhy IS z hlediska různých dimenzí IS.<sup>49</sup> Dílčí architektury tak popisují na jedné úrovni funkce, procesy a data, na další úrovni technologické řešení a na poslední úrovni z toho odvozené softwarové a hardwarové řešení. Jednotlivé dílčí architektury se navzájem prolínají a jedna druhou ovlivňují. Nás bude nejvíce zajímat zejm. model funkční, procesní a technologický. K tvorbě architektonických modelů je možné použít řadu modelovacích nástrojů, zejm. CASE nástroje, které umožňují modely vytvářet vizuální cestou za podpory množství pomůcek.

Při návrhu globální architektury musíme vycházet ze základní funkce IS pro integraci informačních zdrojů, a tím je role zprostředkovatele různých informačních zdrojů a služeb pro potřeby uživatelů. Pro potřebnou úroveň abstrakce se můžeme na takový systém dívat jako na klasický **obchodní model**. Základním cílem uživatelů – **zákazníků** je získat primární informaci, na kterou můžeme pohlížet jako na **zboží**. Producenti a provozovatelé informačních zdrojů vystupují v IS jako **dodavatelé** zboží (primárních informací) nebo poskytovatelé informací o zboží (sekundárních informací). Jednotlivé moduly IS budou tedy obecně tvořit **komunikační rozhraní** pro komunikaci s dodavatelem a zákazníky, moduly pro **evidenci** zboží (organizaci zdrojů), pro **vyhledávání** zboží (vyhledávání zdrojů, sekundárních informací a primárních informací) a pro **získávání** (nákup) a **distribuci** (prodej) zboží (primárních informací) uživatelům. Jednoduše můžeme tento model znázornit takto:



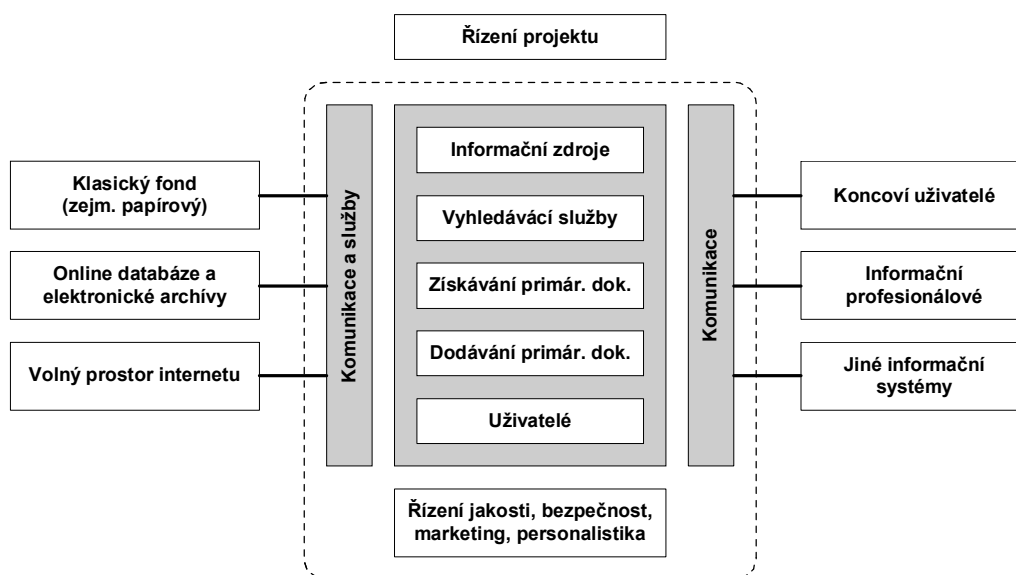
Obr. 8 - Základní návrh globální architektury integračního systému

Při dalším rozpracování dospějeme k podrobnějšímu globálnímu návrhu, který poslouží pro odvození příslušných dílčích architektur. Globální architektura tak slouží jako zastřešující model, ve kterém by se měly všechny dílčí modely střídat a být v souladu. Popsat ho můžeme následovně:

- Blok informačních zdrojů můžeme rozčlenit podle stupně digitalizace a uspořádanosti na **klasický (papírový) fond**, na **elektronické databáze/archivy** a na **volný prostor internetu**.
- V bloku zákazníků je třeba rozlišovat podle různých informačních potřeb **koncové uživatele**, **informační profesionály** a **jiné IS**, kteří využívají služeb projektovaného IS.
- Centrální blok IS se bude skládat z modulů:
  - pro evidenci, správu a podporu **uživatelů**, včetně zajištění autentikace, autorizace a plateb
  - pro akvizici, katalogizaci, klasifikaci a konfiguraci **informačních zdrojů**
  - zajišťující komplexní **vyhledávací služby** (tj. řízení a zpracování dotazů, zpracování výsledků, SDI, linkování atd.)
  - pro **získávání primárních dokumentů**
  - pro bezpečné **dodávání primárních dokumentů** včetně zabezpečení a poplatků



- Všechny zmíněné moduly musí být v IS podpořeny dalšími pomocnými bloky, které musí zajišťovat řízení jakosti, garantovat bezpečnost, zastřešovat procesy, které nelze plně automatizovat a které jsou prováděny lidskou silou, podporovat marketingové procesy atd.



Obr. 9 - Rozpracovaný návrh globální architektury integračního systému

### 2.1.2 Funkční architektura IS pro integraci informačních zdrojů

Budeme-li jednotlivé bloky globálního návrhu dále rozpracovávat a upřesňovat, přejdeme na úroveň návrhu funkční architektury integračního systému. Tento návrh na jednotlivých úrovních postupně rozkládá funkční bloky na dílčí elementární funkce, které odpovídají požadavkům uživatelů. Podrobný návrh funkční architektury tak představuje přesný popis uživatelských a systémových funkcí, kterými bude IS disponovat (vyjadřuje to, co má systém umět). Na základě toho se definují jednotlivé procesy a následně samotný softwarový návrh a s ním spojená softwarová architektura.

Návrh funkční architektury může např. tvořit systematický soupis funkcí s jejich podrobným popisem. Členění funkcí by mělo být hierarchické, funkce by se neměly překrývat a měly by pokrývat všechny situace. Příklad možného zpracování funkčního bloku „Informační zdroje“ ukazuje následující tabulka:

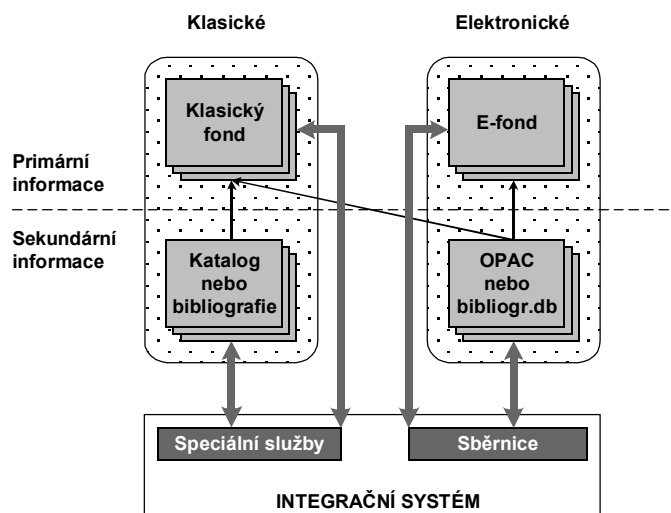
FUNKCE	POPIS A OBSAH FUNKCE
Popis zdroje	- jmenný popis - věcný popis - opatření anotací - kontakty na provozovatele - informace o dostupnosti - možnost exportu a importu
Klasifikace zdroje	- správa klasifikačních systémů - možnost hromadných změn
Technická konfigurace zdroje	- přístupové informace - mapování vyhledávacích atributů - mapování znakových sad - vzdálená autentikace
Statistika využití	- evidence přístupu ke zdroji - evidence počtu stažených objektů
Hlídaní licenčních podmínek	- maximální počet konkurenčních připojení - časová platnost licence

	- odpovědná osoba
Přístupová práva ke zdroji	- povolit/zakázat podle uživatele, skupiny, instituce nebo IP adresy
Životní cyklus zdroje	- akvizice: návrh a schválení nákupu, využívání, obnova licence, ukončení užívání - posílání avíz a výstrah - kontrola platnosti URL a sledování změn obsahu dokumentů

Podoba funkčního návrhu vychází z potřeb konkrétního projektu, příklad možného řešení ukazuje kapitola 5.

### 2.1.3 Procesní architektura IS pro integraci informačních zdrojů

Procesní architektonický návrh vyjadřuje podobu a vztahy jednotlivých procesů IS, které budou zabezpečovat požadované funkce. Již jsme několikrát zmínili, že hlavním úkolem IS pro integraci informačních zdrojů je poskytnout uživateli v jednotném prostředí relevantní **primární** informace pomocí široké nabídky informačních zdrojů. Ke splnění tohoto cíle je třeba navrhnout jednotlivé procesy IS tak, aby dovedly uživatele v různých situacích a v co nejkratším čase k hledanému dokumentu nebo záznamu. Jádrem těchto procesů vychází ze vzájemných vztahů mezi primárními dokumenty a informačními pomůckami, které slouží k jejich evidenci a vyhledávání, a to jak v klasické, tak elektronické podobě (více informací o základních vlastnostech a typech informačních zdrojů – viz kapitola 1.1). Těmto vztahům je třeba do detailu porozumět. Obecně je můžeme znázornit následovně:



Obr. 10 - Procesní vztahy mezi základními skupinami informačních zdrojů

Výše znázorněné vztahy by měly představovat jádro řešení integračního systému pro práci s informačními zdroji. Integrační systém může pomocí datové sběrnice komunikovat s lokálními i vzdálenými elektronickými informačními zdroji, které jsou přístupné online a které obsahují primární a/nebo sekundární informace. Uživatel tak může prohledávat lokální i externí OPACy knihoven, plnotextové databáze článků apod. (pravá polovina obrázku). Elektronické sekundární informace (na obrázku označeno jako OPAC) obsahují přirozenou vazbu na relevantní primární informace v klasické i elektronické podobě, které lze v integračním systému využít online. Uživatel tam může např. z lokálního OPACu vidět klasickou část knižního fondu i sbírky elektronických dokumentů. Klasické sekundární informace (na obrázku označeno jako Katalog) a jejich vazby na klasické primární informace nelze integrovat běžným způsobem (komunikační sběrnici). Online zpřístupnění klasických sekundárních informací lze zajistit postupnou retrokatalogizací/digitalizací nebo jednotlivě pomocí speciálních služeb (na základě dotazu). Klasický fond je dostupný také pomocí speciálních služeb, jako je např. služba elektronického dodávání dokumentů. Speciální služby se vyznačují tím, že v nich vystupuje lidský element,

např. pracovník, který na základě elektronické žádosti vyhledá ve skladu papírový dokument a požadované stránky oskenuje.

**Elektronické sekundární informace** představují katalogové a bibliografické pomůcky, které:

- slouží k získání informací o primárních dokumentech v klasické i elektronické podobě
- zajišťují vazby pro získání elektronických primárních dokumentů (samoobslužně) či pro sestavení žádosti o elektronické dodání klasických dokumentů (s obsluhou)

Mohou být zcela integrovány do IS, protože mohou být k dispozici v reálném čase online bez nutnosti asistence pracovníků knihovny. Připojení se děje na principu přímé datové komunikace přes datovou sběrnici.

**Klasické sekundární informace** jsou fixovány na neelektronická média a nemohou být bez předchozí digitalizace přístupné online v reálném čase. Uživatelé je mohou využít pouze v budově knihovny, kde jsou fyzicky umístěny (lístkové katalogy), nebo mimo IS (tištěné bibliografie). Jinou možností je vzdálené využití prostřednictvím pracovníků knihovny v rámci služeb (telefonický či elektronický dotaz na knihovnu). Z důvodů těchto omezení dochází k hromadné digitalizaci klasických sekundárních informací, zejm. u katalogů knihoven, ať již metodou tradičního skenování (naskenované katalogizační lístky jako obrazy přístupné online) nebo retrokatalogizací (elektronický přepis katalogizačních údajů, manuálně nebo s pomocí OCR).

**Elektronický fond** (přístupný online) je díky různým metodám zpracování, formátů a médií velice různorodý. Může jít o vlastní elektronický fond, který má knihovna popsán a evidován v katalogu, o informace organizované v databázích či elektronických úložištích s alespoň minimálním popisem (strukturou), nebo o volný prostor internetu, který není nijak popsán ani organizován a je ho možno pouze sekvenčně indexovat a následně prohledávat. Knihovna jako profesionální a zkušený katalogizátor zdrojů by se měla snažit zajišťovat a zkvalitňovat evidenční a popisné služby elektronických zdrojů, aby měli uživatelé k dispozici stejný komfort při vyhledávání, na jaký jsou zvyklí u katalogu klasických dokumentů.

**Klasický fond** představuje tradiční oblast dokumentů, kterou knihovny vlastní a spravují. Význam a hodnota klasického fondu je stále nesporná, je však třeba nabízet i odpovídající elektronické služby pro jeho zpřístupnění online. Nabízejí se dvě základní metody:

- průběžná a systematická digitalizace klasického fondu, která ale musí být z finančních a časových důvodů omezena pouze na vzácné a ohrožené dokumenty (metoda „just in case“)
- digitalizace na vyžádání prostřednictvím služeb elektronického dodávání dokumentů (metoda „just in time“)

Z nákresu je zřejmé, že část integračních procesů je možné zcela zautomatizovat a část je třeba zajistit manuálně, protože dochází k manipulaci s fyzickými neelektronickými dokumenty a jejich záznamy. Speciální služby mohou mít mnoho podob a v návrhu integračního systému je třeba s nimi počítat. Speciální služby by měly být napojeny na procesní vstupy a výstupy jednotlivých metod poskytování primárních informací, aby uživatel mohl získat všechny typy dokumentů pokud možno elektronicky. Více informací o službách dodávání primárních informací – viz kapitola 2.8.3.

#### 2.1.4 Technologická architektura IS pro integraci informačních zdrojů

Technologický model IS definuje způsoby zpracování dat, strukturu aplikací IS a uživatelských rozhraní. Zvolené technologické modely nemusí být v celém IS stejné, naopak se často kombinují a doplňují podle potřeby. Podívejme se nyní na několik nejdůležitějších hledisek a pojetí technologického řešení IS.

Velice významným aspektem z hlediska dostupnosti služeb je zvolená **metoda zpracování úloh**. Připomeňme si, že rozeznáváme:

- postupy **interaktivní**, kdy je úloha hned po zadání v reálném čase zpracovávána (např. získání elektronického dokumentu online)
- postupy **řízené událostmi**, kdy je úloha provedena ve stanovený čas nebo při splnění určité podmínky (např. každodenní spuštění SDI nebo expirace platnosti dokumentu)

- postupy **dávkové**, kdy jsou úlohy shromážděny v dávce a hromadně zpracovány (např. rozesílání upomínek nebo zásilek MVS)
- postupy **manuální**, kdy je pro zpracování nutný zásah člověka (např. skenování požadovaného článku z novin pro účely EDD nebo distribuce zásilek poštou)

Postupy zpracování úloh IS se mezi sebou liší zejm. dobou odezvy, což je pro uživatele při využívání služeb často kritickým faktorem. Např. při dodávání primárních dokumentů mohou nastat v zásadě tři situace, které jsou determinovány právě způsobem zpracování požadavku. Uživatel může získat primární dokument buď:

- přímo v systému elektronicky (samoobslužně)  
= **informaci tak získá okamžitě**
- přes EDD jako elektronickou kopii  
= **informaci tak získá do 24 hodin** (v pracovní dny)
- přes MVS jako fyzickou kopii s doručením  
= **informaci tak získá zhruba do dvou týdnů** (v rámci ČR)

Po krátkém čase však již může být informace pro uživatele bezcenná, protože buď obsahově zastaralá nebo již není z hlediska plnění pracovního úkolu, který je často vázán na konkrétní časový termín, pro uživatele aktuální. Informace nedodaná včas ztrácí informační hodnotu.

Architektonické modely můžeme také dělit **podle míry rozčlenění (vrstvení) skupin funkcí**. Typickými skupinami jsou prezentační funkce zajišťující komunikaci s uživatelem (vstupy a výstupy), aplikační funkce zajišťující vlastní logiku zpracování úloh, datové funkce zajišťující manipulaci s daty, distribuční funkce zajišťující obsluhu více uživatelů a provozní funkce zajišťující správu aplikace.<sup>49</sup> Rozlišovat tedy můžeme architekturu:

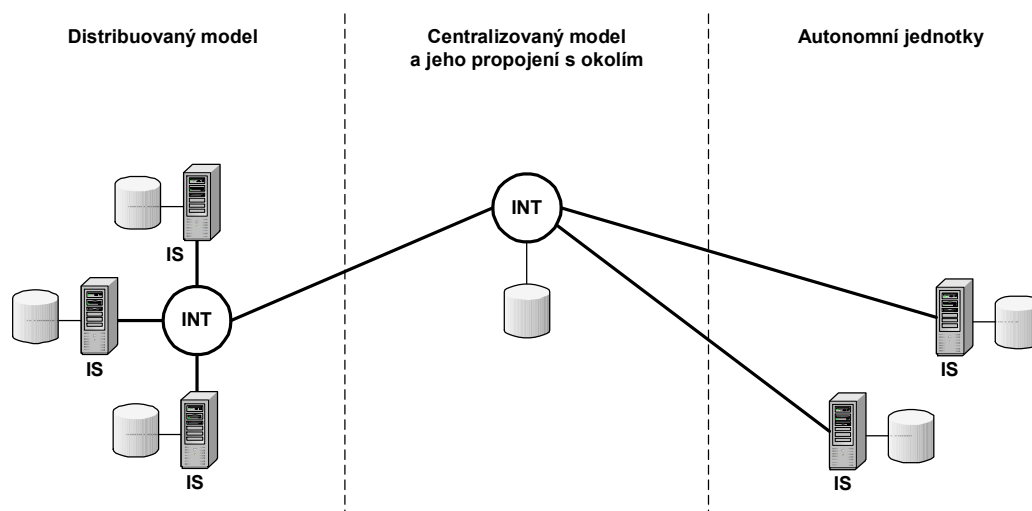
- **monolitickou**, kde jsou všechny funkce přítomny v jednom programu a nejsou vrstvené (příkladem mohou být starší aplikace spustitelné na CD-ROM)
- **dvouvrstvou**, kde je prezentační vrstva oddělena od serveru a je přenesena na klienta, datová a aplikační vrstva však není rozlišena
- **třívrstvou**, kde je oproti dvouvrstvé architektuře navíc oddělena datová a aplikační vrstva na serveru
- **vícevrstvou**, kde jsou vyčleněny ještě další vrstvy, např. distribuční nebo provozní

Monolitická architektura je vhodná pouze pro jednoduché, nenáročné a uzavřené aplikace, což není případ IS pro integraci informačních zdrojů. Z povahy integračního systému, který zajišťuje složité procesy integrace mnoha heterogenních informačních zdrojů pro potřeby mnoha koncových uživatelů, je nanejvýš vhodné použít **architekturu vrstvenou**, kdy je aplikace rozdělena mezi více samostatných funkčních vrstev. Tím je dosaženo dvou základních efektů. Za prvé může dojít v síťovém prostředí k rozdělení aplikace mezi klienty a server (architektura klient/server), čímž je zajištěna nutná nezávislost integračního systému na proměnlivé skupině koncových uživatelů. Za druhé může dojít k oddělení funkčních částí na samotném serveru, čímž je docíleno lepší flexibility, škálovatelnosti a možnosti modifikace za pochodu. Samostatně tak můžeme spravovat a řídit data aplikace, samotné aplikační moduly, uživatelské rozhraní a komunikační moduly.

Integrační systém vystupuje jako server, který vyřizuje požadavky mnoha nezávislých klientů. Server komunikuje s klienty pomocí komunikačního rozhraní, které jako jediná serverová vrstva musí být do jisté míry neměnná, ostatní části serveru se mohou v průběhu provozu systému měnit, aniž by se změny projevovaly u koncových uživatelů (upgrady OS, serverových aplikací, hardware atd.). Klient představuje základní komponentu, která požaduje provedení určité služby, tj. slouží minimálně k zadávání vstupních dat a k zobrazování výsledků. Je tedy alfou i omegou celého integračního systému z pohledu koncového uživatele. Systémoví architekti by měli při návrhu IS vycházet právě z potřeb koncových uživatelů a z technických podmínek, ve kterých koncoví uživatelé pracují.

Při hledání vhodné architektury serverové části řešení pro integrace informačních zdrojů musíme rozhodnout, chceme-li budovat **centralizovaný** nebo **distribuovaný** integrační model. Data a služby můžeme buď fyzicky integrovat do jednoho centrálního prvku nebo můžeme data a služby zprostředkovávat externě ze vzdálených IS. Oba přístupu se navzájem výrazně odlišují a je strategicky důležité vybrat ten správný způsob pro konkrétní případ. Volbu asi nejvíce ovlivňuje

počet a charakter zdrojů, které chceme integrovat. Centralizované a distribuované modely je možné samozřejmě kombinovat a vytvářet **hybridní** řešení. Získáme tak řadu výhod obou modelů a eliminujeme jejich nevýhody. Jedná se často o případy, kdy je integrace realizována několikastupňově (jednotlivé integrační systémy jsou do sebe vnořeny), např. Jednotná informační brána<sup>43</sup> integrující několik fyzických souborných katalogů nebo Google indexující dokumenty z mnoha centrálních úložišť plných textů.



Obr. 11 - Ukázky integračních modelů a jejich propojení do hybridní podoby

### 2.1.4.1 Typy klientů

Respektujeme-li základní požadavek „využívat informační zdroje kdykoli a odkudkoli“, nemůžeme skupinu **koncových uživatelů** zužovat na návštěvníky studoven využívající námi nabízené pracovní stanice. Uživatelé mohou do integračního systému vstupovat ze stolních počítačů s bohatým programovým vybavením, z přenosných zařízení typu PDA či kapesních zařízení jako je inteligentní mobilní telefon. Tato zařízení mohou mít rozmanité operační systémy, velmi rozdílné zobrazovací schopnosti, různé restriktce ohledně připojení na internet či možnosti modifikace nastavení aplikací. Situace je tedy v zásadě zcela odlišná, než z jaké vychází systémové integrace v podnicích, kdy je možné počítačové vybavení zaměstnanců významně ovlivnit, ba dokonce diktovat. Jistě je proto rozumné požadavky na technické vybavení koncového uživatele minimalizovat, čímž bude integrační systém využitelný pro největší skupinu koncových uživatelů. Kromě toho je třeba definovat požadavky na klienty **editorů, informačních profesionálů**, případně **jiných IS**, které budou data a služby integračního systému také využívat.

Podle procesních požadavků můžeme klienty rozdělit takto:

- **tlustý klient** - obsahuje prezentační i aplikační vrstvu, takže je klient schopen vykonávat řadu vlastních funkcí (příkladem může být grafický katalogizační editor AKS pro MS Windows)
- **tenký klient** - obsahuje prezentační vrstvu a žádnou nebo pouze velice omezenou aplikační vrstvu zajišťující např. tisk a vyhledávání na stránce (příkladem může být webový prohlížeč zobrazující pouze HTML/CSS bez aktivní aplikace)
- **terminálový klient** - využívá pouze vstup a výstup koncového zařízení, veškeré zpracování včetně přípravy zobrazení probíhá na terminálovém serveru. Jinými slovy, klient pouze zachytává a na server odesílá vstupy (z klávesnice nebo myši) a zpátky ze serveru pouze přenáší generované obrazovky (příkladem může být Vzdálená plocha Microsoft Windows)

Díky rozmachu internetu, zvýšení minimálních přenosových rychlostí a rozšíření webových technologií můžeme sledovat odklon od technologií tlustých klientů. Pro koncové uživatele, kteří nevystupují jako editoři dat ale jako jejich konzumenti, se již tlustí klienti téměř nikde nepoužívají. Stále se však v některých IS používají pro účely editace dat, kdy je výhodné nabízet bohatší aplikační funkcionalitu již na straně klienta. Hlavní nevýhodou tlustých klientů jsou vyšší

nároky na výkon koncových stanic, problematické nasazení na mobilní zařízení s omezenými operačními schopnostmi, nutnost instalace klienta na koncových stanicích, reinstalace v případě aplikačních změn a nesnadná udržení jednotného nastavení na všech stanicích. Pro správu tlustých klientů ale samozřejmě existuje řada administrativních řešení, které zejm. v lokálních sítích umožňují efektivní správu.

Nejčastěji používaným typem klienta v IS pro koncové uživatele je tenký klient, a to nejčastěji klient webový. To má své důvody. Webové prohlížeče jsou dnes defaultně přítomny ve všech běžných operačních systémech na koncových PC stanicích, na PDA i inteligentních mobilních telefonech (smartphonech) a umí s nimi pracovat největší skupina uživatelů vůbec. Webové prohlížeče splňují s určitými výhradami webové standardy a chovají se tak na všech zařízeních podobně. Podporují také alespoň základní bezpečnostní technologie, např. SSL. Prezentační funkce jsou zajištěny pomocí jazyka HTML/CSS, který navíc zajišťuje i prostředky pro zadávání vstupních dat (webové formuláře). K funkcionalitě webového klienta je třeba kromě hardware s OS již pouze připojení k internetu. Připojení na internet je dnes při práci s informačními zdroji samozřejmostí, otázkou je již pouze rychlost připojení a cena. Můžeme říci, že nejnižší rychlostí, kterou dnes mohou uživatelé přistupovat, je rychlost 48 kbs dosahovaná vytáčeným telefonním připojením nebo čtyřkanálovým GPRS. Ať máme na tuto rychlost jakýkoli názor, je dostatečně vysoká k zajištění rozumného fungování webových aplikací pracujících s textem a střízlivou grafikou. Výhodou webového prohlížeče navíc je, že podporuje technologie, které umožňují do prohlížeče dynamicky nahrávat aplikace, postavené např. na technologii AJAX či Java, čímž může webový klient snadno posloužit i jako tlustý klient.

Při návrhu integračního řešení můžeme samozřejmě narazit i na další požadavky na vybavení klienta, které již samotný webový prohlížeč a OS neřeší a je třeba instalovat další komponenty (např. pro využívání formátu PDF, jehož prohlížeč není v OS Microsoft Windows obsažen). Také pro zajištění zabezpečeného vzdáleného přístupu ke zdrojům mohou být na klienty kladeny další požadavky na nastavení síťového připojení stanice (např. VPN nebo proxy).

Klientskou část integračního systému lze také řešit terminálově, ale to v grafickém režimu klade při větším počtu uživatelů nemalé nároky na výkon terminálového serveru. Terminálová řešení lze nasadit v uzavřených sítích s omezenými skupinami uživatelů, např. v podnikových sítích nebo v knihovnách na koncových stanicích ve studovnách. Výhodou terminálových řešení je garance programového vybavení a snadná centrální správa.

S kvalitou klienta je nedílně spojena problematika uživatelského rozhraní, o které se zmíníme v kapitole 2.2.

#### **2.1.4.2 Centralizovaný model integračního systému**

**Centralizovaný model** IS pro integraci informačních zdrojů předpokládá, že veškerá data a služby budou umístěny na centrálním serveru (nebo skupině serverů tvořících jeden logický celek), ke kterému budou přistupovat klienti. Není přitom rozhodující, o jaká data se jedná - jestli jsou to vlastní primární dokumenty a/nebo popisné záznamy a metadata a/nebo pouze vyhledávací indexy. V praxi to znamená, že všechny uvažované informační zdroje jsou fyzicky centralizovány v jednom IS, kde jsou data uložena, kde probíhá vyhledávání a odkud jsou vráceny výsledky i nabízeny navazující služby. Centralizace je vždy výsledkem dohod připojených účastníků, často s politickou garancí spolupráce. Centralizovaný model může z hlediska tvorby zdrojů fungovat na principu **centrální editace**, kdy jsou data skutečně v reálném čase primárně ukládána do centrální databáze, nebo může pracovat na základě **replikace či synchronizace**, kdy jsou data ukládána do dílčích, lokálních databází a teprve následně přenášena do centrální databáze (k těmto metodám patří i technologie sběru dat OAI). Replikované centrální (označované také souborné) databáze, jejichž data vznikají v jiných lokálních systémech, se potýkají s problémy neaktuálnosti dat a často nejednotnou strukturou vstupních dat, kterou lze řešit pouze přísnými editačními pravidly, validacemi a náročnými konverzemi. Zvláštním případem centrální editace je metoda, kdy jsou data ukládána nejprve do centrální databáze a následně kopírována do lokálních IS - to však není z hlediska centrálního prvku podstatné.

K **výhodám** centralizovaného modelu patří:

- centrálně ukládaná data lze lépe ošetřit proti duplicitě a zajistit jim vyšší kvalitu a integritu
- data mají jednotnou strukturu a jsou jednotně indexována, což se pozitivně projevuje při vyhledávání
- jednoduchá tvorba aplikace integračního řešení
- jednoduchá správa aplikace a dat
- jednotné centrální řešení je investičně průhlednější

K **nevýhodám** centralizovaného modelu patří:

- je nutné zajistit podporu a ochotu k přechodu na centrální model, často pod politickým nebo ekonomickým nátlakem
- je nutné vzdát se (částečně nebo zcela) proprietárních řešení, které mohou kvalitou centrální IS převyšovat
- těžko lze uspokojit často protichůdné potřeby všech účastníků, nutnost provádět kompromisy
- lze aplikovat pouze na uzavřené skupiny a konsorcia
- riziko přetížení centrálního prvku a jeho síťového připojení
- přílišná závislost účastníků na centrálním prvku
- v případě výpadku centrálního prvku dojde k výpadkům IS u všech účastníků

Centralizovaný model je nejvhodnější pro řešení v rámci jedné organizace či konsorcia, kde se účastníci dohodnou na stejných pravidlech a postupech. Příkladem může být např. společnost Ebsco, která nabízí přístup do více než 150 databází a tisíců elektronickým časopisů v jednom systému, nebo centrální knihovně-informační systém nasazený ve fakultních knihovnách jedné univerzity. Méně vhodný je tento model pro nasazení mezi více nezávislých subjektů, zvláště v situaci, kdy již jednotlivé subjekty úspěšně používají svá proprietární řešení. Případy takových projektů z praxe také dobře známe. K těm úspěšnějším patří různé projekty budování fyzických (reálných) souborných katalogů či souborů autorit, kde se i přes určité obtíže daří dosahovat pozitivních výsledků. Naopak příkladem nevhodného nasazení může být pokus o vybudování společného centrálního knihovního systému pro celou Slovenskou republiku pomocí software VTLS Virtua (projekt KIS3G), ve kterém mají být katalogy všech zúčastněných knihoven vytvářeny ve společné databázi pod společnou aplikací. Problémy, které tento projekt provázely, jsou nejlepším důkazem toho, že centralizovaný model nasazovaný mezi velký počet různých subjektů nese řadu kritických problémů a rizik, zejm. v případě, kdy jsou účastníci nuceni ke spolupráci pod politických a ekonomickým nátlakem.

Vezmeme-li v úvahu všechny možné informační zdroje, se kterými může knihovna pracovat (tj. klasický fond, nejrůznější komerční a nekomerční databáze či elektronická archivy a volný prostor internetu), pochopíme, že nelze vytvářet jedinou centralizovanou datovou základnu - globální portfolio informačních zdrojů má nesmírně široký záběr a ohromnou dynamičnost. I kdyby se globálně sjednotily AKS v knihovnách, bude zde stále mnoho dalších komerčních nebo soukromých zdrojů, u kterých není možné vyvinout tlak na sjednocení a na změny. V takovém záběru je jediným možným řešením distribuovaný model integrace.

### 2.1.4.3 Distribuovaný model integračního systému

**Distribuovaný model** IS pro integraci informačních zdrojů se nesnaží informační zdroje centralizovat na jedno místo do jednoho systému, ale ponechává jim autonomii a využívá je na dálku jako služby. Základem toho modelu je síťová komunikace a schopnost distribuovat dotazy a jiné požadavky na služby vzdálených IS. Integrační systém musí dotaz převést do podoby, které bude rozumět vzdálený IS, odeslat ho a následně přijmout a zpracovat vrácené výsledky. Zpracování výsledků běžně obnáší převod do jednotného vnitřního formátu, v případě prohledávání více zdrojů také sloučení s dalšími výsledky a deduplikace, dále řazení podle určitých hledisek apod. Nutnost několikanásobných konverzí dotazů, protokolů a formátů vede ale mnohdy k nepřesnostem a zkreslení ve vyhledávacím procesu. Univerzálnost distribuovaného modelu je určena tím, kolika různými síťovými protokoly a s kolika datovými formáty může pracovat. Vyhledávání ve více informačních zdrojích jedním dotazem lze v distribuovaném modelu zajistit buď **paralelně** (tj. souběžně) nebo **sekvenčně** (tj. postupně). Sekvenční

vyhledávání je technicky jednodušší, ale je problematické z hlediska časových prodlev (více informací o problémech vyhledávání - viz kapitola 2.7).

K **výhodám** distribuovaného modelu patří:

- spolupráce může být zcela volná, stačí poskytnout komunikační rozhraní a přístup k obsahu
- účastníci si mohou ponechat své proprietární systémy
- velká dynamičnost a flexibilita při změnách prvků, účastníci mohou volně přicházet a odcházet
- přirozené rozložení zátěže mezi centrální prvek a dílčí účastnické prvky
- nezávislost celkové funkčnosti na výpadcích jednotlivých prvků

K **nevýhodám** distribuovaného modelu patří:

- nekonzistence v datové i aplikační základně
- velký důraz na konvertory a adaptéry, časté nepřesnosti a zkreslení ve vyhledávání
- náchylnost k duplicitám a proměnlivé kvalitě záznamů a dokumentů
- složitá koordinace nutná k zajištění optimální komunikaci mezi všemi prvky a prosazení společných standardů a pravidel
- komplikované zabezpečení a autentikace

Distribuovaný model je obecně vhodný pro všechny projekty, kde je zřejmé, že nemá význam nebo **nelze dosáhnout sjednocení** nutné pro zavedení centralizovaného modelu. To platí v případech, kde chybí vůle k centralizaci, kde je vhodnější a levnější nasadit distribuovaný model nebo kde chceme integrovat komerční informační zdroje externích provozovatelů, kteří většinou nemají na centralizované integraci s jinými provozovateli zájem a naopak se snaží uživatele motivovat k tomu, aby využívali výhradně jejich nabídku informačních zdrojů. Provozovatelé navíc mohou kdykoli změnit IS nebo dílčí technologie a formáty, což není v centralizovaném modelu možné flexibilně řešit. Distribuovaný model také představuje jediné možné řešení v případech, kde je skupina účastníků neohraničená a proměnlivá a kde dochází k příliš rychlým změnám obsahu, např. v peer-to-peer sítích. Distribuovaný model je dostatečně flexibilní pro nasazení ve všech situacích, i když je toho často dosaženo na úkor kvality vyhledávání. Příkladem mohou být nejrůznější virtuální souborné katalogy, webové metarešeršní systémy využívající více webových vyhledávacích systémů nebo paralelní prohlídače.

V praxi se stále častěji setkáváme s IS, které kombinují centralizované a distribuované přístupy. Jako velmi efektivní metoda se např. ukázalo sklizení metadat o dokumentech přes protokol OAI-PMH a následné budování lokálních vyhledávacích indexů s odkazy na vzdáleně uložené dokumenty.

### 2.1.5 Softwarová architektura IS pro integraci informačních zdrojů

Softwarová architektura vyjadřuje, z jakých softwarových komponent bude IS tvořen a jakým způsobem spolu budou mezi sebou tyto komponenty provázány. Softwarová architektura v sobě odráží požadavky a vlastnosti funkčního, procesního a datového modelu, je také ovlivněna technologickým řešením. Každá komponenta zajišťuje konkrétní funkce, pracuje se vstupními, řídicími a výstupními daty a pro svůj provoz vyžaduje konkrétní operační prostředí.

Při integraci rozptýlených informačních zdrojů využíváme převážně síťovou architekturu, ve které spolu jednotlivé komponenty (vzdálené zdroje, klienti) komunikují přes síťová komunikační rozhraní. Samotný integrační prvek v sobě může zahrnovat několik softwarových architektur podle složitosti, zejm. lineární, kde jsou funkce procesu zpracovávány sekvenčně za sebou, a vrstvenou, kde jsou funkce uspořádány do vrstev a kde mohou funkce vyšší vrstvy využívat funkce vrstev podřízených.<sup>49</sup>

Informační systém pro integraci informačních zdrojů je typický tím, že řada integrovaných softwarových komponent je předem dána jinými výrobci a provozovateli a nelze ji nijak modifikovat. Z dlouhodobého hlediska je samozřejmě možné technické vlastnosti komponent jiných subjektů ovlivnit prosazováním průmyslových standardů, dohod a společných metodik, ale při řešení konkrétních integračních projektů nelze spoléhat pouze na to. V řadě případů musíme na externí zdroje pohlížet jako na **uzavřené komponenty**, které nám poskytují konkrétní služby prostřednictvím provozovatelem určeného komunikačního rozhraní. Takové komponenty



musíme integrovat takové, jaké jsou, a architekturu integračního řešení musíme uzpůsobit tak, aby je bylo možné využít.

Jedním z nejdůležitějších výběrových faktorů při volbě softwarových řešení je **finanční přiměřenost**. Musíme si uvědomit, že finanční náročnost řešení není dána pouze investicemi do vlastního vývoje a implementace, ale také náklady nutnými pro provoz a údržbu a investicemi do dalšího rozvoje. U mnoha komerčních softwarových technologií je nutné platit tzv. roční údržbové poplatky, které mohou v několikaletém horizontu překonat výši původní pořizovací investice, u technologií poskytovaných bezplatně (např. open source) se může objevit množství skrytých nákladů, zejm. za služby, které mohou vlastní provoz výrazně prodražit. Varianty řešení je proto nutné posuzovat z hlediska celkových nákladů na vlastnictví (Total Cost of Ownership) sledujících celý životní cyklus software. Podrobná finanční analýza softwarových řešení je nezbytná.

### 2.1.6 Licenční podmínky využívání informačních zdrojů

Jednou z největších bariér ve využívání informačních zdrojů jsou omezení vycházející z autorských práv a licencí. Při využívání komerčních ale často i nekomerčních informačních zdrojů musí uživatel souhlasit s licenčním ujednáním provozovatele či mít souhlas autora s užitím díla. Nejčastěji se omezení týká různých forem dalšího publikování, kopírování, půjčování, sdílení, uchování, modifikování, prodeje nebo jiné distribuce ať již celku nebo částí dokumentu. Z hlediska využití informačních zdrojů v integračním systému je nejvýznamnější otázka omezení či zákazu kopírování a/nebo sdílení informací a postoj ke sdílení v rámci IS provozovaného třetí stranou. Jinými slovy, **integrovat můžeme pouze takové zdroje a do takové míry, jak nám to umožňují autorská práva a licenční podmínky**. Týká se to zejm. služeb přímého přístupu k elektronickým dokumentům (např. zobrazení plnotextových článků z databáze), elektronického dodávání dokumentů, MVS a budování elektronických archivů.

**Pořizování kopií dokumentů a jejich částí** řeší v ČR autorský zákon č. 121/2000 Sb. a jeho novela 216/2006 Sb. Uživatel si může podle tohoto zákona zhotovit kopii nebo napodobeninu klasických i elektronických dokumentů pro osobní potřebu, ale pouze sám na svém přístroji, např. doma na svém skeneru. V pořizovacích cenách kopírovacích zařízení a nenahraných nosičů jsou již zahrnuty poplatky coby náhradní odměna za pořízení kopií pro osobní potřebu. Knihovny mohou pro osobní potřebu uživatelů zhotovovat tiskové kopie na objednávku bez ohledu na to, dělají-li to za úplaty nebo bezplatně, pokud platí roční poplatek jako souhrn poplatků za každou tiskovou stránku.<sup>4</sup> Pořizování elektronických kopií autorských děl samotnou knihovnou zákon silně omezuje. Kopírování umožňuje pouze v prostorách knihovny, kam se ovšem zahrnuje i lokální počítačová síť, což při některých interpretacích zahrnuje i počítače připojené přes vzdálený přístup do LAN.

**Služby EDD** jsou z hlediska autorského zákona značně diskutabilní a podle některých výkladů tento zákon porušují. Usilovně se proto hledá řešení, zejm. snahou o uzavření kolektivní licenční smlouvy s kolektivním správcem a výhledově nasazením DRM jako účinné ochrany proti zneužívání kopií vytvořené pro účely EDD. Předpokládá se, že kopie budou zatíženy výrazně vyššími poplatky, než na jaké jsme zvyklí dnes. Zákon také umožňuje pořízení tzv. náhodné technické kopie jako nezbytné a nedílné součásti technologického postupu zpřístupnění dokumentu, čehož lze využít např. při elektronickém dodávání dokumentů, kdy je proces uchování a přenosu elektronického dokumentu pouze dočasný a měl by být ukončen vytištěním dokumentu na straně uživatele a zničením elektronické kopie. Požadavek na dočasnost existence kopií a jejich následné zničení mohou dnes zajistit pokročilé technologie, o nichž se zmiňujeme v kapitole 3.3.

Problematickou oblast využívání informačních zdrojů z hlediska licencí a autorských práv představují **online databáze**, které knihovny využívají na dálku a které tedy i v integračním systému figurují jako vzdálené IS třetích stran. Zde mají v praxi nejvyšší váhu licenční podmínky provozovatelů, které definují využití obsahu databází. Velkou část těchto zdrojů provozují zahraniční provozovatelé, a proto často vycházejí z jiných legislativních norem, než které platí v ČR (např. úplný zákaz tisku). Tyto licence často ještě více omezují možnosti využití obsahu než zmiňovaný autorský zákon pro ČR. Při nákupu licencí je vhodné předem ověřit, jakým způsobem lze obsah databází využívat, zejm. z hlediska pořizování kopií a zpřístupňování v jiných IS, kterým je např. integrační systém.

S využíváním obsahu databází v integračním systému souvisí také možnost **přesunu procesu autentikace do jiného IS**, který provozuje jiný subjekt než samotný provozovatel databáze. Ne vždy provozovatel důvěřuje třetí straně při ověřování a má pochopitelné obavy ze zneužití. I zde však mohou pomoci moderní technologie autentikace, zejm. sdílená autentikace (více informací o autentikaci a autorizaci - viz kapitola 2.3).

V rámci návrhu architektury IS pro integraci informačních zdrojů tak knihovny musí zvažovat, jak a jaké informační zdroje mohou integrovat z hlediska důvěryhodné autentikace, z hlediska sdílení obsahu v jiném IS a z hlediska autorských práv. Tyto otázky je vhodné u komerčních zdrojů vyřešit ještě před jejich nákupem, protože lze často k jednomu titulu nebo tematické oblasti nalézt několik alternativních producentů či provozovatelů. Obecně mohou knihovny svým uživatelům nabízet služby poskytující tištěné kopie záznamů či dokumentů (xerox, fax), ale musí za ně odvádět roční poplatky. Mohou také zhotovovat dočasné elektronické kopie pro účely elektronického dodávání dokumentů, ale musí být schopny znemožnit využití daných kopií po uplynutí lhůty potřebné k distribuci na své straně i na straně uživatele, čehož lze technologicky dosáhnout.

### 2.1.7 Zajištění dostupnosti zdrojů a jejich obsahu

Jak jsme již několikrát zmínili, můžeme informační zdroje rozdělit do tří základních skupin: klasický (nejčastěji papírový) fond, elektronické databáze/archivy a volný prostor internetu. Tyto skupiny se mezi sebou liší nejen formou uspořádanosti informací a popisu, ale také **mírou stability obsahu a garantovanou dostupností**.

Z hlediska dostupnosti jsou ve výhodě elektronické formy dokumentů, zejm. ty, které jsou přístupné online, protože se snadno zpřístupňují uživatelům a hlavně je lze sdílet současně v jeden okamžik mezi velkým počtem zájemců, narozdíl od fyzických exemplářů klasických dokumentů, které může mít v jeden okamžik pouze jeden uživatel. Elektronické dokumenty se také snadno zálohují a duplikují, takže je lze snadno chránit. Lze je také lehce indexovat, čímž je zaručeno jejich efektivní vyhledávání. Klasické dokumenty lze vyhledávat pouze s pomocí popisných záznamů a navazujícího vyhledávacího aparátu (např. bibliografický záznam doplněný lokačními značkami umístění ve skladu) a jeho dostupnost může být lehce narušena chybami ve skladovém zařazení či chybné evidenci – to bývá také jednou z příčin vzniku tzv. mrtvého fondu, který nelze vyhledat. Dostupnost fyzického fondu je třeba udržovat pomocí pravidelných revízií fondu a pečlivou evidencí výpůjčních operací. Klasický fond je nutné chránit před destrukcí, před krádežemi atd. a jeho začlenění do automatizovaného IS téměř vždy vyžaduje zapojení manuálních operací (klasické výpůjčky, EDD atd.). Všechny tyto činnosti značně prodražují správu a využívání fyzického fondu.

Elektronická a fyzická fixace informací se liší i ve stabilitě obsahu a vždy přináší pozitiva i negativa současně. Trvale fixovaný obsah sice zaručuje trvalou dostupnost obsahu, ale ztrácí se možnost průběžných aktualizací. Zatímco u klasických dokumentů, kde jsou informace trvale fixovány na záznamové médium, je obsah prakticky neměnný (nebereme-li v úvahu možnost fyzické destrukce), u elektronických forem je situace jiná. U online databází, u dokumentů na internetu, ale často i u elektronických archivů dochází ke **změnám obsahu** (úpravy textů, náhrada obsahů článků pod stejným názvem), k **přesunům či odstranění dokumentů** (změna síťových adres, odstranění souborů z úložiště či záznamu z databáze) nebo k **nedostupnosti z technickým příčin**. Nepříjemným ale častým jevem u elektronických dokumentů na webu je také situace, kdy provozovatelé ukončí svoji činnost (zejm. nezávislí autoři vystavující svá díla a vědecké práce), čímž nenávratně zmizí celá webová sídla. Poměrně vážnou situací je také **ukončení platnosti licencí** pro vstup do komerčních zdrojů, kdy mohou knihovny během jediného dne ztratit přístup do stovek zdrojů a do milionů dílčích dokumentů (např. ukončením vstupu do ProQuestu). Knihovny samozřejmě musí tyto změny sledovat a nějakým způsobem na ně reagovat, aby nedocházelo ke vzniku nekonzistencí a nepřesností v evidencích odkazů a v záznamech. Při trvalé nedostupnosti dokumentů je třeba vyřadit je z evidence, v případě modifikací obsahu je třeba upravovat evidenční záznamy. Tyto kontrolní činnosti nelze při současném objemu elektronických zdrojů provádět manuálně, je nutné implementovat účinné automatické mechanismy nebo je využívat jako externí službu. U všech typů dokumentů je také třeba pravidelně kontrolovat čitelnost obsahu z hlediska poškození média, čitelnosti ve čtecích zařízeních (CD-ROMy) či datových formátech.

### 2.1.7.1 Metody sledování a zajištění dostupnosti

Základní metodou sledování dostupnosti elektronických zdrojů je kontrola platnosti URL adres, autentikace a konektivity. Odkazy či záznamy s neplatnými či duplicitními adresami je nutné zaktualizovat nebo odstranit. Tyto kontroly je možné zcela zautomatizovat, částečně je také možno automatizovat změny či alespoň návrhy změn. Důležitým aspektem je stanovení délky nedostupnosti, po jejímž uplynutí bude zdroj považován za trvale nedostupný nebo nespolehlivý, neboť ke krátkodobým výpadkům v dostupnosti externích zdrojů může docházet prakticky u všech zdrojů. Kontrola tedy nejčastěji vychází z analýzy vrácené hlavičky a kódu odpovědi, jejíž výsledek je pak statisticky zpracován a vyhodnocen z hlediska míry nedostupnosti zdroje v časovém intervalu. Nejčastějšími příčinami nedostupnosti zdroje je změna názvu/adresy serveru, chyby sítě, přesun zdroje na jiné místo v rámci úložiště nebo na jiný server, odstranění zdroje a technické problémy serveru zpřístupňující daný zdroj.

Krátké výpadky v dostupnosti zdrojů lze technicky řešit průběžným ale dočasným cachováním obsahu dokumentů (jako to dělá např. systém Google), tvorbou lokálních indexů vzdálených zdrojů nebo kompletní archivací obsahu. Tyto metody však mnohdy narážejí na otázky autorského práva a licencí, protože jsou považovány za formu pořizování kopií.

Velkým problémem elektronických informačních zdrojů je dnes zajištění dostupnosti z hlediska **čitelnosti** záznamových médií a formátů. Integrovaný systém je ideálním nástrojem pro zajištění zpětné kompatibility starších formátů a protokolů, které nejsou klientem podporovány, a to formou vnitřních konverzí. Uživatel se nemusí starat o to, jestli formát dat, se kterými pracuje, je v originálním zdroji čitelný jeho softwarovým vybavením – integrovaný systém všechna data převede do formátů čitelných v použitém klientovi.

### 2.1.7.2 Metody sledování změn obsahu a zajištění obsahu

Sledování změn obsahu u elektronických dokumentů spočívá v **porovnávání** současného a předchozího obsahu zdrojů pomocí indexace, pomocí popisných či technických metadat, pomocí plnotextového porovnávání apod. Jiné metody je třeba zvolit u samotných plnotextových dokumentů, jiné u strukturovaných záznamů či databází. Zjištěné změny je třeba vždy posoudit z hlediska smyslu a významu, protože některé změny se mohou týkat pouze odlišné formulace textu beze změny významu, jiné mohou význam posunout a jiné mohou představovat úplnou změnu dokumentu (např. text zpravodajských aktualit se může pravidelně obsahově i tématicky měnit). Posuzování změn se většinou provádí intelektuálně, přestože experimentálně vznikají automatické sémantické systémy pokoušející se o rozpoznávání významu textů. Při hlídání změn obsahu informačních zdrojů je vhodné také použít zpětnou vazbu od uživatelů, kteří nás mohou na mnoho nekonzistencí upozornit. Pokud se obsah zdroje změní natolik, že již kvalitou a zaměřením nevyhovuje zvoleným kritériím, je třeba ho vyřadit. Pokud se významově pouze posunul jinam, ale kritériím stále vyhovuje, je třeba provést změny v popisu a evidenci.

Pouhým sledováním změn obsahu samozřejmě nemůžeme ovlivnit fakt, že spousta informací z informačních zdrojů nenávratně mizí. Studie Kongresové knihovny uvádí, že meziročně zmizí přibližně 40% webového obsahu (přičemž zároveň naroste asi o 200% díky publikování nových dokumentů).<sup>27</sup> O uchování těchto zdrojů se snaží různé **archivační a konzervační projekty** (u nás např. WebArchiv realizovaný v Národní knihovně ČR), narážejí však na velká omezení a problémy z hlediska autorského práva a licencí. Vážným problémem archivace těchto zdrojů je také zajištění formátů a technologií pro budoucí čitelnost zdrojů, protože textové i grafické formáty a komunikační protokoly se rychle mění. V každém případě zapojení elektronických archivů do integrovaného řešení má ohromný přínos, protože přináší spoustu informací, které již v reálném prostředí volného internetu a databází zmizely.

### 2.1.8 Manuálně prováděné služby jako náhrada nedostatečné automatizace

Ani v době pokročilé automatizace v knihovnách nemůžeme systémově nahradit všechny služby nebo všechny jejich složky strojovou prací. Jedná se zejm. o činnosti, kdy je nutné manipulovat s fyzickými předměty (např. tištěnými svazky časopisů) nebo kdy je nezbytná intelektuální práce člověka (zejm. práce s přirozeným jazykem, obsahová analýza apod.). Některé tyto činnosti nelze strojově nahradit vůbec, zejm. z důvodu garance kvality, nebo proto, že by strojové řešení bylo neúměrně nákladné. Je věcí strategické volby v každém konkrétním projektu, které činnosti budou automatizovány a které nikoli. Důvodem k automatizaci by měla být vysoká frekvence

rutinních operací a zároveň dostupnost finančně přiměřeného a spolehlivého strojového řešení pro danou operaci. Spolehlivost strojového řešení je klíčová hlavně u intelektuálních činností a často není příliš uspokojivá. Pokud je ale frekvence činnosti již vysoká natolik, že není možné ji personálně zvládnout, vyplatí se nasadit automat i při vědomí nižší kvality. Příkladem může být např. automatické indexování plných textů v objemu, který přesahuje ekonomicko-personální možnosti dané instituce.

Zapojujeme-li do integračního systému služby prováděné zcela nebo částečně manuálně, je třeba dodržet určité zásady, aby nebyl narušen plynulý tok procesů IS. Obecně by mělo platit, že IS by měl uzpůsoben tak, aby s ním mohla obsluha pracovat efektivně a nemusela se nepřírozeně přizpůsobovat systému. **Výstupy** IS určené pro obsluhu by měly být pro člověka dobře čitelné a jednoznačné, **vstupy** určené pro zadávání výsledků manuální operace by měly být lehké zadatelné a odolné proti chybám. Jinými slovy, IS by měl plně respektovat lidské vlastnosti a omezení. Ve fázi implementace je samozřejmě nezbytné obsluhu manuálních částí IS dostatečně **proškolit** a následně průběžně kontrolovat kvalitu jejich práce, protože lidský faktor způsobuje značnou variabilitu v jakosti a výkonu.

Nejčastějšími operacemi, které jsou vykonávány manuálně, jsou: skenování dokumentů pro účely EDD, popis zdrojů, věcná analýza a indexování, manipulace a odesílání fyzických dokumentů poštou pro účely MVS, vyřizování dotazů uživatelů v rámci služby Ptejte se knihovny apod.

## 2.2 UŽIVATELSKÉ ROZHRAŇÍ

Základním cílem při budování informačních systémů pro integraci informačních zdrojů je poskytnout uživateli uživatelské rozhraní, ve kterém bude moci efektivně a příjemně pracovat se všemi informačními zdroji a využívat rozmanité služby. S metodikou tvorby kvalitního uživatelského rozhraní je spjato mnoho různých názorových směrů a technických postupů. Rozhraní je však vždy definováno **daty**, se kterými pracuje, **operacemi**, které s daty provádí, a **jazykem**, kterým komunikuje s uživatelem. Cílem této práce není definovat jediný správný způsob řešení uživatelského rozhraní, poukážeme ale na několik klíčových momentů při návrhu uživatelského rozhraní z hlediska integrace heterogenních prvků. Základem je, že rozhraní musí být intuitivní, jednoduché, nezastrašující, lehké pochopitelné a musí podporovat rychlost práce. Uživatel by neměl mít při práci tušení o systémové architektuře aplikačních a datových vrstev systému, jednak z důvodu bezpečnosti a jednak kvůli přívětivosti rozhraní.

Uživatelské rozhraní svým řešením vytváří tzv. jazyk rozhraní, tedy konkrétní metody a nástroje, které může uživatel v IS použít při plnění svých úkolů. Nástroje by měly uživateli umožnit realizovat požadavky pomocí minimálního počtu kroků. Použití nástrojů musí být dobře ošetřeno proti výjimkám, aby nedocházelo k nepředvídaným situacím a chování IS. Příliš přísná pravidla pro využívání pracovních nástrojů může uživatele omezovat v práci, naopak příliš volné pojetí může vést ke zmatkům a chybám. Při návrhu uživatelského rozhraní je nutné udělat **podrobnou analýzu potřeb a zvyklostí uživatelů**, k čemuž lze např. použít analýzy a statistiky využití jiných podobných řešení nebo betatesting prototypu IS ve spolupráci s koncovými uživateli. Podle ambicí a zaměření projektu lze vytvářet různá uživatelské rozhraní pro různé typy uživatelů, umožňovat uživatelům měnit nastavení a obsah komponent rozhraní apod. Největší komfort práce by mělo přinášet uživatelské rozhraní v portálech zajišťujících vysoký stupeň interaktivity a modifikovatelnosti. Samozřejmě všech uživatelských rozhraní by měla být podrobná online nápověda a dokumentace ve formě uživatelské příručky.

Řešení uživatelského rozhraní může být výrazně ovlivněno zvolenou technologií klienta, která může možnosti rozhraní limitovat. Nejrozšířenějším a nejoblíbenějším typem klienta pro koncové uživatele je bezesporu klient webový, jak jsme zmiňovali v kapitole 2.1.4.1. Při dodržení platných standardů při tvorbě webových aplikací, zejm. W3C definicí pro HTML/CSS a JavaScript/DOM, a při rozumném omezení aktivních komponent v Javě a Flash, lze dosáhnout maximální možné univerzálnosti pro prohlížení obsahu i tisk.

Pravidly pro návrh vhodného uživatelského rozhraní se zabývá mnoho iniciativ, ať již na úrovni komerčních společností (Microsoft, Apple Macintosh atd.) nebo neziskových organizací. Pro vývoj

webových aplikací, na které jsou IS pro integraci informačních zdrojů většinou orientovány, je nejvýznamnější iniciativa **WAI** (Web Accessibility Initiative) organizace W3C.

Když odhlédneme od otázek grafického designu, který spadá do oblasti estetického, mnohdy silně subjektivního vnímání, můžeme u uživatelského rozhraní integračního systému hodnotit zejm.:

- schopnost uživatele orientovat se v rozhraní
- možnosti navigace
- možnosti vyhledávání
- míru konsolidace prezentace různorodých dat pocházejících z různých informačních zdrojů
- rychlost odezvy

### 2.2.1 Orientace uživatele v rozhraní IS

Zvláště u rozsáhlejších IS je důležité, aby se uživatel v prostředí snadno zorientoval a rychle vyhledal potřebné informace nebo službu. Mluvíme o tzv. **čitelnosti rozhraní**. Obecně platí, že orientace je snadná, když jsou příbuzné objekty zobrazovány jednotně. To se týká zobrazování plných textů, prezentace databázových objektů, webových stránek, nabídek menu apod. Uživatel si lehkou navykne na konkrétní vzhled daných objektů a následně může pracovat rychleji, jistěji a s menším množstvím navigačních a vyhledávacích chyb. Texty a různé výpisy by měly být krátké nebo rozdělené na několik částí, popř. prezentovány jako náhledy. Měly by být také maximálně přehledné a dobře strukturované. Na začátek dokumentů, zejm. článků, je také vhodné umístit krátké vyjádření hlavního momentu, tzv. čelo článku. Významnou roli hrají také umístění na obrazovce, použitá písma, barvy a použité grafické prvky. Obecným pravidlem by měla být střídmost, jednoduchost a účelnost, což se pozitivně projeví i v univerzálnosti pro různá koncová zařízení.

### 2.2.2 Možnosti navigace

Uživatel by měl mít k dispozici přehledný systém ovládacích prvků a příkazů, pomocí kterých může ovládat rozhraní IS a pohybovat se v něm. Ovládací prvky by měly být logicky rozčleněny podle funkcí a nejpoužívanější z nich by měl mít uživatel stále po ruce. Ovládací prvky a příkazy by měly mít ve všech situacích stejnou funkci a nemělo by docházet příliš často k situacím, kdy je nelze použít. K dispozici by také měla stále být kontextová nápověda. Navigace může být účinná pouze tehdy, je-li dodrženo jednotné uspořádání a rozložení dílčích prvků rozhraní. Je také vhodné použít způsob navigace, chování a názvy ovládacích prvků, na která je uživatel zvyklý z jiných systémů, zejm. operačních systémů a kancelářských aplikací. Důležité je, aby prezentované výsledky vyhledávání ve zdrojích byly dostatečně sjednoceny, aby je mohl uživatel procházet jednotným způsobem.

### 2.2.3 Možnosti vyhledávání

Uživatel může mít na vyhledávací služby různé nároky, je proto vhodné nabídnout několik úrovní vyhledávání od jednoduchého vyhledávání typu Google až po propracované vyhledávací rozhraní pro zadávání složitých a přesných dotazů. Vyhledávací rozhraní by mělo být opatřeno kvalitní nápovědou, případně průvodci, kteří uživatele vizuálně provedou procesem formulace dotazu a zbaví tak uživatele nutnosti učit se systémový dotazovací jazyk. Více informací o vyhledávání přináší kapitola 2.7.

### 2.2.4 Konsolidace výsledků

Při integraci heterogenních informačních zdrojů přicházejí z různých zdrojů různé výsledky v různých formátech. Na výstupu v uživatelském rozhraní by měly být tyto výsledky v maximální míře sjednoceny, např. konverzí do jednotného strukturovaného formátu, převodem do definovaného zobrazení apod., aby uživatel pracoval s podobnými prezentacemi dat. Jednotlivé informační zdroje mohou mít ale natolik odlišnou strukturu dat, že jsou možnosti sjednocování omezené. Sjednocování záznamů však není jediné možnost, jak výsledky konsolidovat. Často přináší daleko vyšší přidanou hodnotu budování vyšších informačních entit, jejichž vlastnosti pocházejí z různých informačních zdrojů. Například můžeme na výstupu dynamicky vytvářet záznam knihy, který se bude skládat z bibliografické informace (pocházející z bibliografické

databáze), recenze na knihu (z knihkupecké databáze), obálky knihy (ze samostatné databáze naskenovaných obalů) a odkazem na další díla daného autora (generovaného fasetovým způsobem vedlejším vyhledáváním z bibliografické databáze). Více informací o vyhledávání a zpracování výsledků přináší kapitola 2.7.

### 2.3 AUTENTIKACE A AUTORIZACE

IS pro integraci informačních zdrojů často přinášejí obsah, který je třeba zpřístupnit pouze oprávněným uživatelům. Toho lze dosáhnout pouze spolehlivou autentikací a autorizací. Problematika autentikace a autorizace (AA) v IS je natolik široká, že představuje zcela samostatné téma. Pro účely této práce pouze shrneme základní rysy a poukážeme na aspekty, které ovlivňují architekturu IS pro integraci informačních zdrojů.

V odborné literatuře se můžeme setkat se třemi různými formami výrazu pro označení procesu ověření identity uživatele: autentizace, autentikace a autentifikace. Každý z nich přišel z jiné jazykové oblasti. Autentizace byla přejata z německého výrazu Authentisierung, autentikace z anglického authentication a autentifikace z francouzského authentication. Jejich význam je v češtině zcela synonymní. V této práci budeme používat variantu **autentikace**.

**Autentikace** je proces, ve kterém IS ověřuje identitu uživatele nebo služby. Po úspěšné autentikaci uživatele IS ověří, ke kterým zdrojům a službám má uživatel oprávnění, a zajistí mu odpovídající přístupy – tento proces se nazývá **autorizace**.<sup>14</sup> Autentikace je tak předpokladem uskutečnění autorizace. Díky AA může IS uživatelům poskytovat přístup ke zdrojům podle oprávnění, k personalizovaným službám, jako jsou uživatelské schránky, kam si uživatelé mohou ukládat výsledky své práce, možnost uchovávání nastavení pracovního prostředí či identifikačních údajů. Uživatelé účty také mohou být napojeny na platební a kreditní systém umožňující mikroplatby za služby apod. Spolehlivá a bezpečná autentikace je v IS klíčově důležitá.

To, čím se výrazně liší úloha autentikace a autorizace IS pro integraci informačních zdrojů od jiných autonomních IS, je potřeba sdílet tyto mechanismy mezi více subjekty, zejm. se vzdálenými IS sloužícími jako informační zdroje. Z předešlých kapitol víme, že takové zdroje jsou velmi různorodé a používají často velmi odlišné způsoby autentikace a autorizace. Proto je v integračním systému kladen velký důraz na integraci dílčích autentikačních a autorizačních mechanismů, kterým obecně říkáme **single sign-on** (SSO) - uživatel se tak nemusí hlásit do každého zdroje samostatně a pamatovat si množství hesel či jiných přístupových kódů – autentikuje se pouze jednou v integračním systému, který ověření a oprávnění poskytuje dál. Integrovaná AA je také předpokladem paralelního vyhledávání. Procesy AA mohou být integrovány buď v **centrálním** modelu, ve kterém používají všechny služby a aplikace jeden společný systém AA, nebo v modelu **distribuovaném** (federativním), ve kterém je proces AA prováděn na vzdáleném systému AA, který se na základě ověření za konkrétního uživatele zaručí a potvrdí jeho oprávnění volané službě. Ve federativním modelu volaná služba nezíská všechny osobní informace uživatele, ale pouze atributy nutné ke garanci oprávnění, čímž je dosaženo maximální ochrany uživatelů a celkové bezpečnosti.

Autentikační integrace může být dosaženo zejm. poskytováním ID a hesel k informačním zdrojům pro konkrétní uživatele nebo skupiny z centrálního úložiště, vystupováním pod jedním centrálním uživatelem, využíváním externích databází uživatelů (např. ověřováním přes adresářové služby) a jejich certifikátů, evidencí rozsahů IP adres příslušných institucí nebo používáním krátkodobých proxy certifikátů. Autorizační integrace může být založena na výměně LDAP atributů (role uživatele nebo členství v uživatelských skupinách) z adresářových služeb či atributových certifikátů z atributových serverů nebo u webových zdrojů na technologii typu Shibboleth. Vlastní oprávnění jsou určena definovanými politikami uživatelů a skupin. Předpokladem účinné autorizace jsou spolehlivě **zabezpečená** autentikace a proti zneužití zabezpečený celý integrační systém. V opačném případě by autorizace ztrácela smysl. Více informací o bezpečnosti přináší kapitola 2.4.

Použité autentikační a autorizační metody ovlivňují celkovou architekturu integračního systému a rozhodují o spolehlivosti a důvěryhodnosti celého systému. Pokud má provozovatel informačního zdroje souhlasit a důvěřovat vzdálené autentikaci a autorizaci, musí mít jistotu, že je integrační řešení schopno tento úkol **bezpečně** zajistit, a často požaduje průběžnou **evidenci a statistiky přístupu** z dílčích počítačových stanic, protože k provozovateli jsou již často veškeré přístupy směřovány centrálně bez možnosti rozlišení konkrétních uživatelů. Ne vždy je však u dílčích zdrojů možné AA integrovat, ať již z technických nebo licenčních důvodů (model SSO je vždy založen na vzájemné **důvěře** dílčích aplikací).

Jak již bylo řečeno, téměř všichni komerční provozovatelé online informačních zdrojů i velká řada těch nekomerčních chrání přístup ke svým aplikacím a informacím nějakou formou zabezpečení proti neoprávněnému přístupu. Nejčastěji se jedná o některý z těchto autentikačních metod:

- zadáním ID uživatele a hesla, která jsou uživateli staticky přidělena
- zadáním jednorázového hesla, které je při každé autentikaci vygenerováno a zasláno uživateli např. jako SMS
- přístupem z IP adresy, která je u provozovatele registrována jako oprávněná
- prokázáním identity pomocí digitálního podpisu nebo jiného kódu (v softwarové nebo hardwarové podobě)
- použitím biometrických metod

Výše uvedené metody se často kombinují a jsou vyžadovány současně, protože tím stoupá stupeň bezpečnosti. V takovém případě hovoříme o vícefaktorové autentikaci. Příkladem může být použití chipové karty (1. bezpečnostní faktor), u které je třeba navíc zadat PIN (2. bezpečnostní faktor).

Autentikace **statickým heslem** je nejběžnější, ale nese s sebou rizika zneužití hesla (heslo je možné fyzicky i elektronicky odposlechnout nebo uhodnout). Je proto vhodné komunikaci alespoň v okamžiku autentikaci zabezpečit (např. zapouzdřením nebo šifrováním) a zavést účinnou politiku správy hesel (např. omezenou dobou platnosti hesel, jejich minimální délkou atd.).

Používání **jednorázových hesel** snižuje možnost zneužití, protože heslo je při každém přihlašování jiné. Jednorázové heslo vygeneruje IS na základě ID a zašle ho uživateli např. e-mailem nebo jako SMS. Z důvodu možného odposlechu hesla je opět vhodné použít zabezpečenou komunikaci.

Autentikace na základě **IP adresy** počítače nebo přenosného zařízení má výhodu, že uživatel nemusí ručně zadávat ID a heslo. Majitelem IP adresy je vždy konkrétní subjekt, který může být zároveň uživatelem licencí informačního zdroje, proto je tato autentikační metoda oblíbená u komerčních provozovatelů informačních zdrojů. Nevýhodou je nepřenositelnost, takže uživatelé nemohou bez dalších technologií používat zdroje mimo prostory dané instituce. Řešením jsou technologie zabezpečeného vzdáleného přístupu.

Osobní **digitální podpis** uživatele může sloužit jako autentikační kód do IS. Nejčastěji má podobu softwarového certifikátu. Je poměrně bezpečný, ale nevýhodou je, že ho uživatel musí mít pro tyto účely stále po ruce. Používání digitálních podpisů nebo jiných kódů usnadňují hardwarová řešení, kdy se uživatel podepisuje např. vložením chipové karty s uloženým podpisem do čtecího zařízení.

**Biometrické metody** autentikace řeší problém jednoznačné identifikace uživatele. Uživatel si nemusí pamatovat žádná hesla, nemusí s sebou nosit žádné chipové karty ani certifikáty, protože jako vstupní data slouží určité parametry jeho vlastního těla. Běžně se dnes používají metody snímání otisků prstů, oční duhovky nebo sítnice, celé ruky či dlaně, snímání obličeje a analýza hlasu. Biometrické metody ale narážejí na etické problémy a na možnosti politického, sociálního a ekonomického zneužití.

Vyžaduje-li provozovatel vzdáleného informačního zdroje, aby uživatel přistupoval z předem zaregistrovaných IP adres, musí knihovna řešit vážný problém. Při nákupu licence musí knihovna uvést provozovateli **rozsah IP adres svých stanic**, ze kterých bude oprávněna se zdrojem pracovat. K informačnímu zdroji se tak uživatelé dostanou pouze z počítačů knihovny, externě jim bude přístup odepřen. Uživatelé ale dnes chtějí služby knihoven využívat vzdáleně, nejčastěji z domova a ze zaměstnání. Navíc často používají ke komunikaci mobilní zařízení, zejm. notebooky, jejichž IP adresa se při cestování mění, nebo tito uživatelé při cestách používají různá stacionární zařízení, podle toho, kde se zrovna nacházejí. Knihovna, která chce vyjít uživatelům vstříc, tak musí umožnit přístup těžko ohraničitelné a velmi dynamické skupině koncových zařízení, což představuje poměrně náročný úkol, hlavně z hlediska bezpečnosti.

Řešením jsou nejrůznější technologie zajišťující **vzdálený přístup** do lokální sítě nebo **propůjčující IP adresy** instituce vzdáleně (více informací o vzdáleném přístupu - viz kapitola 3.4).

## 2.4 BEZPEČNOST

Dnes již nikdo nepochybuje o tom, že aspekty bezpečnosti jsou jedněmi z nejdůležitějších otázek efektivity a úspěšnosti projektů IS. Přechodem aplikací do síťového a často distribuovaného prostředí jsou možnosti selhání a napadení různých komponent skutečně veliké. Bezpečnost IS představuje samostatnou, mezioborovou disciplínu. Pro účely této práce ukážeme pouze základní aspekty bezpečnosti, které nejvíce ovlivňují celkovou architekturu integračního systému.

Na bezpečnost IS se musíme dívat obecně ze dvou pohledů. Jednak z pohledu **ochrany dat před zneužitím** a jednak z pohledu **stability a důvěryhodnosti** IS. IS pro integraci informačních zdrojů vystupuje jako zprostředkovatel externích informačních zdrojů koncovým uživatelům a pracuje tedy s nedůvěryhodnými komponentami, jejichž chování nemůže přímo ovlivnit. S externími zdroji musí komunikovat, aby mohl nabízet jejich obsah, ale tato komunikace představuje potenciální bezpečnostní hrozbu, kterou je třeba eliminovat. Integrační systém musí zároveň chránit obsah externích zdrojů před zneužitím v souladu s licenčními podmínkami. Tradiční hrozbu samozřejmě představují také koncoví klienti, kteří do systému vstupují. Předmětem ochrany a zabezpečení v integračním systému tedy jsou tyto základní části:

- **Zprostředkovávaná data a služby externích provozovatelů informačních zdrojů**  
Data a služby provozovatelů dílčích informačních zdrojů je nutné chránit před zneužitím podle platných licenčních podmínek, dohod o provozu a podle platných autorských práv. Zvláště u komerčních informačních zdrojů je tato oblast citlivá, protože se často jedná o licence v hodnotě statisíců korun. Ochrana spočívá ve spolehlivé a bezpečné autentikaci, chráněné distribuci komerčních informací a v nasazení technologií proti nelegálnímu kopírování.
- **Součásti a data vlastního integračního řešení**  
Ochrana komponent vlastního integračního systému většinou není citlivá na zneužití dat, ale nestabilita a časté výpadky mohou způsobit ztrátu důvěry uživatelů i provozovatelů integrovaných zdrojů v integrační systém. Provoz vlastního integračního systému může být ohrožen selháním hardware a software, výpadky napájení nebo poruchami sítě, vinou obsluhy, napadením ze strany klientů ale i vzdálených zdrojů, zavirováním apod. Ochrana by se tedy měla soustředit především na účinnou správu přístupových práv, zabezpečení komunikace, na zálohování dat a efektivní plány obnovy systému v případě výpadku.
- **Uchovávaná data koncových uživatelů**  
Data spojená s identifikací a činností koncových uživatelů spadají do oblasti osobních informací a nakládání s nimi se musí řídit dikcí zákona. Zákon omezuje, resp. definuje možnosti shromažďování těchto dat, zpracovávání a povinnost jejich zabezpečení proti zneužití. Integrační systém musí být navržen tak, aby těmto legislativním požadavkům vyhověl a zejm. zajistil účinnou ochranu uložených i komunikovaných dat.

Již jsme zmínili, že v integračním řešení jsou přítomny i komponenty, za jejichž funkčnost, stabilitu a bezpečnost odpovídají jiné subjekty. Jsou to na jedné straně koncová zařízení klientů a na druhé straně všechny služby a data využívaná z externích systémů (tj. samotné informační zdroje, adresářové služby apod.) Těchto komponent může být veliké množství a mohou být velice různorodé, proto i projevy nestability v těchto komponentech mohou nabývat mnoho různých podob (výpadky konektivity, chyby databázových systémů, nehlášené změny ve vyhledávání, nefunkčnost v době zálohování apod.). Pokud v těchto částech dochází k výpadkům, projeví se problémy v našem integračním systému, aniž bychom za to nesli odpovědnost a mohli to nějak ovlivnit. Je proto důležité ve všech situacích celkových i dílčích výpadků komponent integračního systému **informovat uživatele**, o jaký problém se jedná, co je jeho příčinou a **kdo za problém nese odpovědnost**. Jedině tak může být integrační systém uživateli vnímán jako spolehlivý a řízený.



Oblasti zabezpečení každé výše uvedené části můžeme zjednodušeně rozdělit podle funkčních vrstev IS takto:

- **Autentikační vrstva**  
Autentikaci můžeme v první řadě zabezpečit šifrovanou komunikací **SSL** založenou na principu PKI certifikátů. SSL je široce podporovaný standard, který podporují všechny prohlížeče na všech běžných OS, jeho nasazení by nemělo být problematické. Jinou alternativou je použití jiné šifrované komunikace, např. použití **VPN** pro zajištění vzdáleného přístupu, kdy proces autentikace do integračního systému probíhá také zabezpečeně. S VPN souvisí i další forma ochrany autentikace a tou je omezení přístupu pouze pro **povolené rozsahy IP adres**. Můžeme také použít speciální autentikační protokoly, jako je např. Kerberos nebo RADIUS. Zmíněné zabezpečení lze různě kombinovat a dosáhnout vyšší účinnosti.
- **Komunikační vrstva**  
Po ověření uživatele je nutné dále zabezpečovat komunikaci a to v průběhu celé uživatelské relace, protože uživatel může pracovat s osobními údaji a placenými a licenčně chráněnými informacemi či dokumenty. I zde se jako jednoduché ale poměrně účinné řešení nabízí **SSL**. Na straně serveru je třeba zabezpečit všechna komunikační rozhraní, tedy zejm. servery zajišťující datovou komunikaci HTTP či Z39.50, externí adresářové služby LDAP a různé typy univerzálních aplikačních rozhraní API, bez ohledu na to, komunikují-li s koncovým uživatelem nebo obsluhujícím serverem. Pro komunikaci postavenou na nestavovém protokolu HTTP je třeba navíc zvolit bezpečný způsob udržování ověřené relace. Také by měly být ošetřeny všechny možné vstupy od uživatele.
- **Aplikační a datová vrstva**  
Na úrovni ochrany vlastní aplikační a datové vrstvy integračního systému je třeba udržovat bezpečné a aktuální verze OS a aplikací, účinně se bránit proti škodlivému software a mít spolehlivý systém zálohování. Klíčově důležitý aspekt bezpečnosti je existence efektivních havarijních plánů, podle kterých je možné v případě selhání bezpečnostních opatření v co nejkratší době obnovit důležitá data a funkce integračního systému. Předpokladem bezpečnosti je také precizní správa přístupových práv uživatelů a skupin k souborům a adresářům na serveru, zejm. ze strany komunikačních rozhraní (webserver). Samotná uložená data, pokud jsou citlivá, je vhodné šifrovat, ať již v databázi nebo v souborovém systému.

Důležitou součástí bezpečnosti je nepřetržité sledování komunikace a procesů. Velký význam proto mají systémové protokoly (logy) a statistiky, které by se měly podrobovat průběžné analýze. Před zahájením ostrého provozu a poté průběžně v pravidelných intervalech by měl být také prováděn bezpečnostní audit.

S bezpečností souvisí i zabezpečení neautomatizovaných složek IS, např. ochrana knižního fondu, se kterým se musí manipulovat při realizaci služeb EDD. Ochrana fondu se snaží zajistit dostupnost obsahu dokumentů v čase, tj. konzervovat klasický fond a archivovat ten digitální. Na tato ochranná opatření by měl brát integrační systém ohled.

## 2.5 AKVIZICE INFORMAČNÍCH ZDROJŮ

Integrace informačních zdrojů má značný vliv na budování fondů, tedy na akvizici zdrojů. Klasický způsob akvizice (jako proces **soustavného a průběžného získávání** informačních zdrojů pro danou instituci) lze při integraci centralizovat nebo alespoň koordinovat tak, aby zapojené instituce vzájemně pokrývaly co nejširší záběr zdrojů bez zbytečných duplicít pomocí **sdílení zdrojů**. To umožňuje nejen optimálně využít finančních prostředků na nákup, ale hlavně rozdělit pokrytí konkrétních oborů podle zaměření nebo podle vzájemných dohod mezi institucemi. Uživatel integračního systému pak snadno nalezne potřebné dokumenty vlastněné jakoukoli z institucí a pomocí přímého zobrazení, EDD nebo jiných služeb je pohodlně získá.

Zajímavou alternativou k průběžné akvizici (označované jako Just In Case) jsou moderní směry akvizice, při které jsou dílčí dokumenty získávány až **v okamžiku potřeby uživatele** (Just In Time). Dokument, např. článek, knihovna zakoupí či jinak obstará až na základě požadavku uživatele. Myšlenka to není nová, nicméně až dnes, díky moderním technologiím zajišťujícím potřebnou rychlost vyřízení požadavku, může být tento typ akvizice efektivně nasazen. Takový

typ nákupů dnes podporuje celá řada provozovatelů zejm. článkových databází (Thomson Gale, Ingenta Connect, Anopress a další), zároveň ho zajišťují samotné knihovny formou svých služeb sdílení zdrojů (meziknihovní výpůjčka, EDD v rámci skupiny spolupracujících knihoven atd.). Díky této možnosti jsou knihovny postaveny před kritické rozhodnutí: mají kupovat celé ročníky časopisů nebo si mají pouze zajistit popisné informace o článcích a plné texty článků dokupovat až na vyžádání? Je známým faktem, že výtěžnost elektronických časopisů je malá, řadu článků si nikdo nikdy nepřečte, proto je efektivnější kupovat pouze individuálně požadované články. **Je překvapivé, že české knihovny nakupují drahé zahraniční článkové databáze, přestože jejich výtěžnost ze strany uživatelů je mizivá a pohybuje se v rámci stovek či dokonce pouhých desítek článků ročně.** Předplatné článkových databází se vyplatí pouze při velkém využití, kde se vysoká cena paušálu rozpočítá do velkého počtu využitých článků. K tomu je třeba mít potřebný informační aparát, jako databáze referátů, recenzí či záznamů s anotacemi, aby si mohli uživatelé udělat představu o uvažovaném článku ještě před objednáním. Stejný princip platí i pro další typy dokumentů, knihy, audio soubory nebo i slovníková hesla, která lze získat na požádání bez nutnosti kupovat celou encyklopedii nebo příručku. Integrace těchto nákupních zdrojů s pomocnými databázemi může proces nákupu značně zautomatizovat, čímž je efektivita dodávání na požádání výrazně zvýšena. V rámci integračního systému mohou být navíc zajištěny licence i přístupy k nákupním zdrojům jednotně pro celé konsorcium.

Pro posouzení výhodnosti nasazení JIT akvizice je třeba znát celkové obchodní podmínky konkrétního dodavatele. Např. u Ingenta Connect, který má asi největší záběr časopiseckých titulů pro EDD na světě (cca 24 000 titulů), je třeba složit minimální kredit 500 USD, ze kterého následně knihovna čerpá poplatky za dodání jednotlivých článků. Roční poplatek za využívání služby činí 100 USD. Cena za jednotlivé články dodávané v elektronické podobě se pohybuje v rozmezí 20 až 45 USD.

Je třeba pečlivě evidovat, kolik článků uživatelé ročně využívají, následně je pak možné zhodnotit, zda je výhodnější roční paušální předplatné využívání databáze (JIC) nebo zda-li je ekonomicky výhodnější platit za konkrétní vyžádané články.

Akvizice může získávat některé zdroje bezplatně, některé musí kupovat v rámci rozpočtů nebo dotací. Je-li zdroj poskytován bezplatně, nemělo by to automaticky znamenat, že bude do integračního systému připojen. Každý integrovaný zdroj by měl splňovat kritéria obsahové kvality, technické způsobilosti a musí odpovídat **přísné akviziční politice**. Při integraci agregátorů informačních zdrojů do nadřazeného IS vzniká závažný problém **překrývání dílčích informačních zdrojů**.<sup>3</sup> Např. dvě článkové databáze, které svým záběrem pokrývají stovky časopiseckých titulů, mohou mít značný překryv, čímž budeme mít řadu časopisů ve své virtuální sbírce duplikovaně.\* Akvizice by proto měla průběžně analyzovat překrytí celé kolekce informačních zdrojů knihovny a snažit se optimalizovat portfolio zdrojů tak, aby byla míra překrytí co možná nejmenší.

Název databáze	Všechny tituly	Jedinečné tituly	Tituly v překryvu
ProQuest 5000	7 372	2 534	4 838
EBSCO MasterFILE	10 618	3 330	7 288
Ingenta Connect	5 928	1 553	4 375
<b>CELKEM</b>	<b>23 918</b>	<b>7 417</b>	<b>16 501</b>

Obr. 12 - Příklad analýzy překryvu časopiseckých databází

\* Při paralelním prohledávání zdrojů s obsahovým překryvem je nutné provádět deduplikaci a slučování nalezených informačních objektů.

Nezbytným procesem akvizice také musí vždycky být **zpětná vazba** informující o využití informačních zdrojů, aby knihovna věděla, co se kupuje zbytečně a o co je naopak zvýšený zájem. Zpětnou vazbu využití zdrojů zajišťující procesy sledování a zpracování statistik přístupů, viz kapitola 2.9.3. Informace o zájmech a potřebách uživatelů by měly zajišťovat pravidelně prováděné průzkumy zaměřené na oborové pokrytí, jazykové pokrytí apod. Je až s podivem, jak velké množství knihoven nemá přehled o jazykových schopnostech svých uživatelů a o jejich profesních a studijních zájmech. Důležitým hlediskem při zjišťování uživatelských potřeb by měly být také **náměty** pro akvizici (dezideráta), ať již vytvářené samotnými koncovými uživateli, kteří jsou často schopni zhodnotit význam zdroje pro daný vědní obor lépe než akviziční pracovníci, nebo generovanými na základě analýzy neúspěšných vyhledávacích dotazů. Všechny procesy zajišťující zpětné vazby o uživatelských potřebách lze automatizovat např. sběrem a následným zpracováním formulářů a jejich integrací do IS.

**Akviziční politika** se přímo odráží v integračním procesu: určuje, jaké zdroje budou zapojeny a v jakém rozsahu, komu budou určeny a za jakých podmínek. Akviziční politika zajišťuje zejm.:

- ▶ vymezení cílové skupiny uživatelů
- ▶ vymezení struktury fondu
- ▶ vymezení rozsahu fondu podle typů informačních zdrojů
- ▶ definuje dílčí rozpočty na pořizování informačních zdrojů
- ▶ určuje pravidla sdílení informačních zdrojů z pohledu akvizice

Specifickým znakem akvizice informačních zdrojů pro integrovaný IS je skutečnost, že pořizuje informační **zdroje velmi rozdílných typů**, které se budou cílově integrovat v jednom systému a navzájem se budou doplňovat. Jedná se o klasický (zejm. papírový) fond, elektronické databáze či archivy a volný prostor internetu. Zatímco v akvizici klasického fondu se knihovny většinou dobře orientují, u databází a zejm. volného prostoru internetu často nevědí, jakou akviziční politiku zvolit. Elektronické zdroje jsou z hlediska využití flexibilnější, ale neměly by představovat dražší alternativu k tištěným zdrojům, alespoň ne pro uživatele (kdy by bylo využití klasického fondu bezplatné a vstup do databází placený), a neměly by být pro uživatele hůře dosažitelné, např. kvůli velkým požadavkům na technickou zdatnost uživatele. Je věcí strategického rozhodnutí, jestli knihovna koupí tištěný časopis, který po nákupu již natrvalo zůstane v jejím fondu ale který může být hůře využitelný, nebo jeho elektronickou verzi v rámci roční licence, který ale v okamžiku ukončení licence natrvalo ztratí. Rozumným řešením se může zdát již zmíněný nákup článků na vyžádání, kdy knihovna časopis předem vůbec nekupuje. Zcela specifický přístup vyžadují zdroje volného prostoru internetu, který pro knihovny představuje bezesporu největší výzvu ze všech tří zmíněných oblastí. Knihovny mají historickou příležitost nabídnout uživatelům předvybrané, kvalitní a ověřené internetové zdroje, přehledně popsané, klasifikované a organizované podle různých kritérií v přehledné struktuře IS. Mohou tak nabídnout nenahraditelnou přidanou hodnotu, které nemůže při současných technických možnostech nabídnout žádný automat či indexovací robot. Akvizice internetových zdrojů je velmi náročná, vyžaduje nové postupy a zkušenosti, má velké nároky na rychlost a musí ještě těsněji spolupracovat s katalogizační a technickou složkou IS (zejm. v otázkách zajištění dostupnosti zdrojů a jejich obsahu – viz kapitola 2.1.7). Protože u velké části volného internetu, kterou nelze automaticky indexovat, nepomohou vyhledávací stroje typu Google, a navíc vyhledávání často nepřináší na prvních pozicích ty nejkvalitnější zdroje, je při akvizici internetových zdrojů nutné plně využít námětů a tipů od uživatelů, od odborníků a pedagogů, sledovat odkazy v jiných dokumentech, na konferencích či diskusních skupinách apod. Alternativou je také zakoupení kvalitního komerčního adresáře internetových zdrojů, který přináší kvalitní popis a garanci platnosti odkazů (např. Gale Web Feet).

Vážný problém akvizice představuje již zmíněná ohromná **nestálost webových zdrojů**, kdy se velká část webového obsahu mění nebo úplně ztrácí. Tuto situaci se snaží řešit velké konzervační projekty (v ČR např. projekt WebArchiv), které vybrané webové dokumenty archivují a zajišťují tak jejich perzistenci. Rychlý technologický vývoj záznamových médií a čtecích zařízení (tedy zastarávání) však tyto ambice značně komplikují, spolu s omezeními podle dikce autorského zákona a licencí.

Akviziční procesy je vhodné zahrnout do celkového **řízení životního cyklu informačních zdrojů**, které sledující zdroje od okamžiku jejich vytipování a doporučení k akvizici, schválení, nákupu či pořízení, zapojení do integračního systému, využívání, prodlužování licencí až v konečné fázi jejich vyřazení. Pro tyto řídicí procesy se používá označení **ERM** (Electronic resource management) a o jejich mezinárodní standardizaci se snaží zejm. iniciativa DLF ERMI (Digital Library Federation Electronic Resource Management Initiative). Jejím cílem je nabídnout producentům elektronických informačních zdrojů a vývojářům ERM systémů standardy, které zajistí automatickou výměnu platebních, licenčních, přístupových, technických a statistických informací mezi producenty a jejich zákazníky (což jsou z našeho pohledu zejm. knihovny). Komunikované informace tak mohou automaticky ovlivňovat stav životního cyklu daných informačních zdrojů v IS, čímž je dosaženo nejen vysoké efektivity ale hlavně vysoké míry organizace.

Všechny tyto aspekty akvizičních procesů je třeba při integraci informačních zdrojů důkladně zvážit.

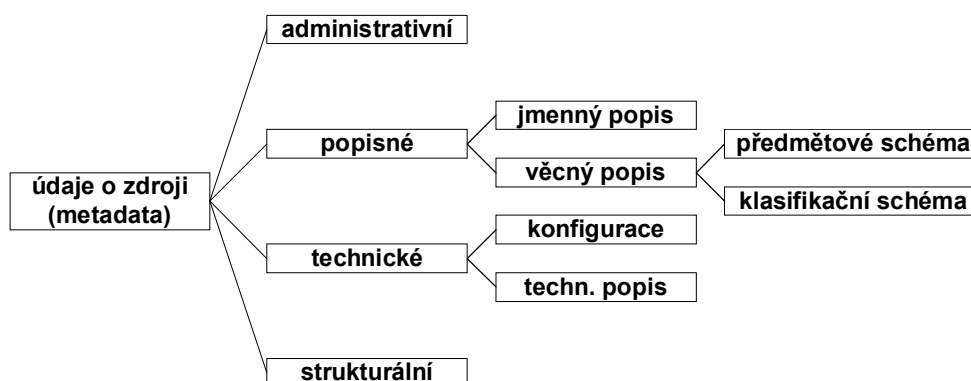
## 2.6 POPIS A ORGANIZACE INFORMAČNÍCH ZDROJŮ

Primární dokumenty můžeme prohledávat buď přímo plnotextově nebo prostřednictvím popisných záznamů (např. bibliografických). Popisné záznamy umožňují vyhledávat dokumenty strukturovaně podle jednotlivých popisných polí, což může přinést větší sémantickou přesnost než vyhledávání plnotextové a při větších objemech dat i vyšší rychlost. U klasických dokumentů je to i jediný způsob vyhledávání. U netextových objektů (např. nasnímaných obrazů) navíc popisné záznamy umožňují vyhledávat podle textových údajů a integrovat takové sbírky s ostatními textovými zdroji. Stejně jako popisujeme dílčí primární dokumenty, je třeba v integračním systému popsat všechny informační zdroje, které do systému zapojujeme, aby je mohl uživatel efektivně nalézt a následně použít. Vyhledávání však není jediný důvod, proč informační zdroje popisujeme. Popis a organizace informačních zdrojů v integračním systému slouží k následujícím účelům:

- **evidence** zdrojů na jednom místě, aby uživatelé viděli kompletní informační portfolio organizace (co organizace nabízí)
- **věcný popis** zdrojů, aby uživatelé věděli, co jaký zdroj obsahuje, jakých se týká oborů, jaké je jeho pokrytí apod.
- **vyhledávání** zdrojů podle různých kritérií (využití polí jmenného popisu, předmětového zařazení, zařazení do klasifikačního systému apod. k vyhledávání)
- **technický popis** přinášející údaje o technických nárocích na využití, o specifikaci formátů atd., zároveň může sloužit pro účely konfigurace vyhledávání v integračním systému
- napojení na **řízení životního cyklu** zdroje po celou dobu jeho využívání v organizaci

### 2.6.1 Popis informačních zdrojů

Údaje, které jsou u jednotlivých informačních zdrojů evidovány, můžeme přehledně rozdělit následovně:



Obr. 13 - Základní členění údajů o informačním zdroji

**Administrativní** metadata slouží k evidenci informací o licencích, řízení životního cyklu zdroje, řízení přístupu ke zdroji atd. **Popisná** metadata slouží k vyhledávání zdroje a jeho popisu pro koncové uživatele. **Technická** metadata popisují technické vlastnosti zdrojů, zejm. z hlediska formátů a protokolů, zároveň slouží k jejich konfiguraci v integračním IS. **Strukturální** metadata vyjadřují vztahy s příbuznými zdroji (vytvářejí např. kolekce).

Výběr vhodných standardů pro jmenný a věcný popis je věcí každého konkrétního projektu. Je třeba zvážit hledisko rozšířenosti a budoucnosti standardu a prognózu dalšího vývoje integračního řešení, aby daná popisná schémata vyhovovala i při výrazném nárůstu počtu zdrojů a rozšíření oborového záběru. Z hlediska možné spolupráce s jinými integračními systémy je vhodné přiklonit se k standardům, které se při spolupráci nejčastěji používají. Pro úplnost pouze připomeňme, že pro jmenný popis informačních zdrojů jsou nejpoužívanějšími současnými standardy MARC, DC či různé formy RDF/XML, pro věcný popis potom klasifikační schémata MDT, DDT, LC, Konspekt či různá oborově zaměřená schémata typu MeSH. Jako jednoznačné identifikátory popisovaných digitálních objektů se nejčastěji používají perzistentní URL, URN, DOI, SICI a ISBN/ISSN. Díky rozvoji technologií tzv. sémantického webu, se začínají objevovat sémanticky strukturované dokumenty, z nichž lze v rámci vymezené oblasti generovat příslušné ontologie. Ontologie se tak mohou v budoucnu stát nástupci klasických tezaurů, které s velkou mírou přesnosti vystihují koncepty a vzájemné vztahy popisovaných objektů.

Z popisu informačního zdroje by měl uživatel v první řadě okamžitě poznat, jaké informace zdroj obsahuje a jak se dostane k primárním informacím, pokud je zdroj neobsahuje přímo. Informační zdroje by měly být přehledně členěny podle oborů, podle typů dokumentů, podle míry vnoření (viz kapitola 2.6.2), podle možnosti získat primární dokument a podle licencí.

Z hlediska integrace informačních zdrojů bychom měli dobře zvážit, jakým způsobem budeme v popisu řešit problémy vícejazyčnosti (integrujeme zdroje české i zahraniční) a jazykové flexe, které představují nejvýznamnější bariéry při vyhledávání a lokaci zdrojů. Nejčastějším řešením **vícejazyčnosti** při vyhledávání jsou zavedení souběžných vícejazyčných popisných polí (např. název zdroje v originálu a zároveň vždy v češtině) a/nebo použití klasifikačních systémů nezávislých na jazyku. **Jazykovou flexi** nejčastěji řešíme používáním formalizovaných výrazů v základních tvarech a/nebo implementací lexikálních systémů pro vyhledávání podle slovního základu (v Hradci Králové, Královéhradecký --> Hradec Králové).

### 2.6.2 Popis různých typů dokumentů a integrace

Katalogizační standardy rozlišují různé typy elementárních primárních dokumentů, které vyžadují specifický popis (monografie, článek, webová stránka atd.). Přestože se tyto dokumenty navzájem liší, je jejich sjednocení v rámci integrace poměrně snadné, protože mají stejný vztah k primárním informacím - přímo je obsahují. Problém nastává v okamžiku, kdy se do integračního systému zapojují typy zdrojů, které vystupují jako soubory elementárních dokumentů nebo informací (časopisy, sborníky, databáze atd.). Mluvíme o tzv. **vnořování informací**.

Současné metody popisu (katalogizace), které se odrážejí v možnostech vyhledávání, nedokáží bohužel plně reagovat na situaci, kdy jsou elektronické informační zdroje do sebe navzájem vnořovány. Integrační procesy, při nichž dochází k agregacím zdrojů do nadřazených celků, vytvářejí nové typy informačních zdrojů, se kterými původní katalogizační standardy nepočítaly a na které nyní reagují zcela nedostatečně. Dokonce ani nové katalogizační iniciativy se nedokáží s tímto problémem vypořádat. Zatím tedy nezbyvá, než vnořování při vyhledávání řešit v integračních systémech proprietárně, aby uživatelé získali ze všech zdrojů srovnatelné výsledky.

Příkladem vnořování může být plnotextový článek (= dokument), který je součástí elektronického časopisu (= soubor dokumentů), který je součástí databáze časopisů (= databáze jako agregátor), která je zahrnuta v informačním portálu instituce (= soubor databází). Jedná se tedy o tři úrovně vnořování, kdy na každé z nich musí být popis článku odlišný, podle způsobu přístupu k danému článku:

<článek>  
<článek> IN: <časopis>

<článek> IN: <časopis> IN: <databáze>  
 <článek> IN: <časopis> IN: <databáze> IN: <portál>

Narozdíl od tištěných článků, může elektronický článek vystupovat na několika úrovních současně, což způsobuje vážné problémy při jedinečné identifikaci. Tento problém není pouze formální, ale má praktické dopady na vyhledávání a prezentaci vrácených dat. V důsledku klíčově ovlivňuje možnosti získání primárního dokumentu. Uživatelé jsou v integračních systémech velmi zmateni, když se jim při vyhledávání některých zdrojů vrací přímo primární dokumenty, v jiných zdrojích získávají primární dokumenty až na základě dalšího vyhledávání a z jiných zdrojů získávají pouze sekundární informace sloužící jako informační pomůcky. V řadě konkrétních projektů, zejm. různých informačních portálů, se ukázalo, že tento problém je jedním z dominantních překážek při získávání informací.

Jednotlivá vnoření můžeme chápat jako mezivrstvy, které je výhodné při integraci zcela odbourat a pracovat s primárními dokumenty přímo. Toho je možné dosáhnout samozřejmě pouze elektronickou cestou. Často to úspěšně řeší provozovatelé článkových databází tak, že umožňují prohledávání článků napříč mnoha časopiseckými tituly. Nad rámec jednoho vydavatele je však nutné tento problém řešit individuálně, pokud to technické a licenční podmínky dovolí. Řešením může být budování samostatných indexů metadat s přímými odkazy na primární objekty (harvesting metadat), vyhledávání primárních objektů bez účasti proprietárních rozhraní (např. přes Z39.50) atd.

### 2.6.3 Možnosti popisu podle stupně digitalizace a uspořádanosti

Na začátku této kapitoly jsme zmínili, že informační zdroje můžeme podle stupně digitalizace a uspořádanosti rozdělit do tří základních skupin: na klasický (zejm. papírový) fond, na databáze a elektronické archivy a na volný prostor internetu. Každá tato skupina vyžaduje od knihoven provozující integrační systém jiné přístupy k popisu.

**Klasický fond** představuje běžnou oblast zájmu knihoven, která je již kvalitně zpracovávána. Podařilo se úspěšně zavést sjednocující katalogizační pravidla a formáty, vznikají také nové návrhy a iniciativy reagující na aktuální problémy. Výsledkem popisu klasického fondu jsou katalogy OPAC a bibliografické databáze, tedy sekundární informace, dnes již převážně v elektronické formě. Problém vnořování je zde poměrně úspěšně řešen (popis článku v rámci časopisu nebo sborníku). I zde však existují problémy v popisu a následném vyhledávání, např. při evidenci rozpětí let odebíraného periodika, kde záznam obsahuje např. údaj „1987-2002“ - pro efektivní vyhledávání by musel IS při indexování pracovat s celým datumovým intervalem, nikoli plnotextově s dvěma uvedenými roky.

**Elektronické databáze a archivy** představují informační zdroje, které slouží jako organizované soubory dílčích elektronických primárních a/nebo sekundárních dokumentů/informací. Protože tato část informačních zdrojů nabývá v knihovnách stále většího významu a objemu, zejm. u časopiseckých článků a bibliografií (např. nákupem licence k databázi Ebsco knihovna získává více než 3000 časopisů, což je mnohonásobně více, než kolik odebírá tištěných časopiseckých titulů), je třeba se jejím efektivním popisem zabývat. Bohužel velký problém způsobuje zmíněné vnořování, které je zde často i několikanásobné. Z tohoto pohledu pro popis databází a e-archivů dodnes neexistují uspokojivé standardy. Jistou nadějí představují iniciativy jako např. METS definující standardy pro popis objektů digitálních knihoven, ale je otázkou delšího času, jak se osvědčí a rozšíří.

Popis elektronických databází/archivů a jejich obsahu na úrovni dílčích elektronických dokumentů by měl být zcela **integrován s popisem klasických dokumentů**, aby uživatelé na jednom místě našli nejen exempláře z klasického fondu ale i z fondu elektronického. Bohužel tomu tak ve většině knihoven není, což se neblaze podepisuje na nízkém využití elektronických zdrojů, protože je uživatelé nenaleznou. Takové evidence mohou knihovny vytvářet samy nebo je mohou přebírat jako službu, kdy se o údržbu stará externí poskytovatel. Běžně se jedná o záznamy elektronických časopisů, ke kterým má knihovna přístup, o nakoupené bibliografické databáze, referátové databáze, vybrané a popsané webové dokumenty apod.

Pokud je v integračním systému žádoucí zpřístupnit některé informační zdroje (dílčí dokumenty) přímo, např. jako nabídku elektronických knih v rámci konkrétní předmětové kategorie, je často výhodnější udělat to prostřednictvím popisných záznamů a databáze. Tím sice vznikne zdánlivě zbytečná mezivrstva, ale docílíme tak sjednocení všech informačních zdrojů na stejnou úroveň vyhledávání. Elektronické knihy tak budeme moci vyhledávat stejně jako knihy z klasického fondu v katalogu, stejně jako články v článkových databázích apod. Stejným způsobem můžeme zpracovávat webové dokumenty z volného prostoru internetu.

**Volný prostor internetu** je oblast, ke které se knihovny staví často velmi rozpačitě. V kapitole o akvizici jsme již zmínili, že knihovny mohou v této oblasti nabídnout nenahraditelné služby s vysokou přidanou hodnotou. Jednou z těchto hodnot je právě precizní a přehledný popis vybraných internetových zdrojů. Uživatelé tak mohou k jednotlivým vědním oborům získat seznam prověřených a kvalitních informačních zdrojů, které by jinak velmi těžko hledali pomocí internetových vyhledávačů či internetových katalogů. Ke jmennému a věcnému popisu volných internetových zdrojů je pro uživatele velmi přínosné zhotovovat krátké anotační poznámky či hodnocení, osvědčil se i systém sdílených poznámek uživatelů (folksonomie), který lze zpětně využít při vyhledávání jako jedno z hledisek.

**Nedostatky internetových vyhledávačů a katalogů** představují šance pro knihovny:

- ▶ Internetové prohledávače mapují pouze indexovatelnou část webu, kam se dostanou roboti, navíc s časovou prodlevou. Mnoho internetových dokumentů je navíc proti návštěvám robotů ošetřeno. Nejaktuálnější, málo citované nebo chráněné dokumenty zůstávají pro většinu uživatelů internetu skryty. Pořadí a relevance výsledků vyhledávání jsou často zneužívány (např. tzv. Google bomb).
- ▶ Internetové katalogy jsou nejednotné a živelně členěny, popis zdrojů je nespolehlivý a neseriózní, protože není vytvářen nezávisle. Pořadí a předstíraná relevance zdrojů v kategoriích je často vytvářena na komerčním základě „předplacené pozice“. Řada autorů svoje dokumenty v internetových katalozích neregistruje.

#### 2.6.4 Organizace informačních zdrojů

Organizace informačních zdrojů v informačním systému představuje způsob, jakým jsou zdroje nabízeny a prezentovány uživateli. Integrační systém může obsahovat řadu cenných zdrojů, ale pokud je uživatel není schopen v systému jednoduše najít a využít, je jejich přínos nulový. Z praxe známe řadu takových negativních případů: knihovny utrácejí vysoké finanční částky za nákupy licencí, ale výtěžnost zdrojů je nízká. Integrační systém by měl organizovat informační zdroje tak, aby uživatel viděl celou nabídku přehledně a byl schopen intuitivně nebo s pomocí průvodců zdroje lehce využívat. I zde musíme při návrhu integračního systému stále pamatovat na náš základní cíl: dodat uživateli relevantní primární informace.

K přehledné organizaci zdrojů slouží v první řadě vhodně implementované **klasifikační schéma**. Pomocí klasifikačních skupin mohou uživatelé přehledně nalézt relevantní informační zdroje podle vědního oboru nebo jiného předmětového hlediska. Výhodou klasifikace je její nezávislost na přirozeném jazyku, takže dobře poslouží pro členění českých i zahraničních zdrojů a může být lehce lokalizována do dalších jazyků rozhraní. Rozčlenění zdrojů podle klasifikačních tříd je také základem při budování tzv. oborových informačních bran, které jsou specializovány na konkrétní vědní obory, čímž je zvýšena relevance a specializace nabízených zdrojů.

Druhým základním nástrojem při výběru zdrojů v IS je jejich **vyhledávání**, při kterém je využíván popisný záznam (jmenný i věcný) každého zdroje. Uživatelé tak mohou zdroje vyhledávat typicky podle názvu zdroje, klíčových slov, předmětových hesel, typu informačního zdroje (plný text, katalog, bibliografie, digitální sbírka atd.), jako je tomu zvykem u záznamů dokumentů.

Významnou organizační pomůckou pro uživatele je možnost vytvářet si **vlastní sestavy oblíbených zdrojů**, ke kterým se mohou uživatelé při pravidelné práci vracet.

Samostatnou kapitolou je pak oblast zveřejňování a propagování nových přírůstků a novinek ve fondech. Velké možnosti zde představuje pravidelný odběr nových informací pomocí kanálů **RSS**. RSS kanály je vybaveno stále více webových systémů a sídel, proto je lze začít v integračních IS

globálně využívat. V integračním systému můžeme odchyťovat RSS kanály a následně je zobrazovat v podobě aktualit z vybraných zdrojů, např. nových přírůstků, nebo je dokonce archivovat a následně v nich vyhledávat. Pokud bude definice RSS součástí technických metadat internetových zdrojů v integračním systému, mohou si uživatelé vytvářet vlastní výběry a získávat pravidelně informace o změnách v oblíbených zdrojích, podobně jako tomu je u služeb SDI.

## 2.7 PROBLEMATIKA VYHLEDÁVÁNÍ V INTEGRAČNÍCH SYSTÉMECH

Při návrhu architektury vyhledávacích služeb v integračním systému bychom měli, jak již bylo řečeno, vždy sledovat základní cíl integrace informačních zdrojů: **dodat uživateli primární dokument**. Jak jsme již zmínili v kapitole 2.1, na cestě k primární dokumentu musí uživatel nejprve nalézt relevantní informační zdroj(e) pro vyhledávání a zajistit si k nim přístup a oprávnění, následně musí v těchto zdrojích vyhledat informace o primárních dokumentech a nakonec na základě těchto informací musí vyhledat vlastní primární dokumenty. Na tento vyhledávací proces musí bezprostředně navazovat služby zajišťující zobrazení nebo dodání primárních dokumentů uživateli.

Nalezení relevantních informačních zdrojů je spjato s problematikou organizace a evidence zdrojů, kterou popisuje kapitola 2.6.

Při vyhledávání vlastních informací rozlišujeme čtyři základní typy vyhledávání, které můžeme v integračním systému podporovat:

- vyhledávání podle **popisu** (podle bibliografických informací nebo jiných metadat)
- vyhledávání podle **citací** (podle citačních vazeb mezi dokumenty/autory)
- **plnotextové** vyhledávání
- **procházení** (browsing, např. abecední výpis apod.)

Dodání primárního dokumentu je zajištěno buď již vlastním vyhledáváním informací nebo na základě linkování či navazujících služeb typu EDD, viz kapitola 2.8.

Vyhledávání informací je rychle a dynamicky se rozvíjející oblast IT, která technologicky směřuje k využívání znalostí a umělé inteligence. Řada experimentálních projektů a iniciativ vytváří vyhledávací modely založené na dobývání znalostí z dat a textů, s výrazným zaměřením na web jako na „nekonečnou“ množinu dokumentů. K nejaktivnějším patří projekty mezinárodního konsorcia WIC (Web Intelligence Consortium) a laboratorní experimenty společnosti Google. Technologický posun v této oblasti nutně ovlivňuje zájem koncových uživatelů o služby knihoven, proto je třeba tyto trendy neustále sledovat.

Problematiku vyhledávání v integračních systémech můžeme rozdělit do tří základních okruhů: na problematiku vlastních vyhledávacích technik a indexování, na problematiku dotazovacích jazyků a na problematiku zpracování výsledků vyhledávání. Následující kapitoly o těchto okruzích v základu pohovoří. Odráží se v nich všechny problémy dříve zmíněné v tomto oddílu, protože ve vyhledávacím procesu se projevují všechny nedostatky připojených informačních zdrojů a jejich způsobu integrace.

### 2.7.1 Vyhledávací proces v integračním systému

Samotný proces vyhledávání v informačním systému, který integruje mnoho informačních zdrojů, je realizován ve dvou krocích:

- **výběr informačních zdrojů**, ve kterých chceme vyhledávat
- **vlastní vyhledávání** informačních objektů

Podle zkušeností je to pro koncové uživatele nejhůře pochopitelná vlastnost IS a uživatelé nejsou často schopni v systému intuitivně pracovat. Je úkolem uživatelského rozhraní a návrhem jeho funkcí se s tímto problémem co nejlépe vypořádat. Fáze výběru informačních zdrojů může být např. eliminována předdefinovanou sadou nejvýznamnějších zdrojů či možností prohledat všechny zdroje z nabídky.



Zmínili jsme již, že jsou do sebe informační zdroje často vzájemně vnořovány, takže na výstupu vyhledávání není koncový dokument. Výsledkem může být také popis jiného informačního zdroje, ve kterém je třeba dále hledat. Tento problém musí řešit sofistikovaný způsob prezentace vrácených výsledků, z nichž některé ukazujeme v kapitole 2.7.7.

### 2.7.1.1 Vyhledávání podle stupně centralizace

Při projektování IS pro integraci informačních zdrojů je klíčové, bude-li vyhledávání informačních zdrojů pojato **centrálně** či **distribuovaně**. Centralizovaný model počítá s tím, že bude budována centrální databáze všech dílčích informačních zdrojů nebo alespoň jejich centrální vyhledávací index. Příkladem centrální databáze jsou např. fyzické souborné katalogy, příkladem centrálně budovaných indexů jinak vzdálených informačních zdrojů je např. systém Google nebo systémy založené na sběru metadat pomocí protokolu OAI-PMH. Budování centrálních úložišť můžeme označit jako princip JIC (just in case), protože data jsou v integračním systému připravena dříve, než jsou skutečně potřeba. Distribuované vyhledávání naopak využívá vyhledávací procesy na vzdálených IS, kterým pošle dotaz a přijme od nich výsledek vyhledávání, příkladem mohou být systémy založené na protokolu Z39.50. Integrační systém tak vzdálená data využívá až v okamžiku, kdy zpracovává dotaz – tento princip označujeme jako JIT (just in time).

Existuje spousta příznivců i odpůrců obou technologických směrů a je třeba vždy zvážit případ od případu, jaký způsob bude pro daný projekt vhodnější. Centrální model přináší samozřejmě výhody při samotném vyhledávání, protože záznamy a selekční údaje v centrální bázi lze optimalizovat, normalizovat či obohacovat a prohledávat celou databázi současně bez nutnosti nasazování prostředků pro paralelní vyhledávání, ale jeho budování a zejm. údržba centrálního úložiště je velmi náročná. Nejčastějším kritériem volby mezi modely je celkový objem dat, finanční náročnost řešení, síťová konektivita a její spolehlivost, ochota provozovatelů dílčích informačních zdrojů spolupracovat a autorská práva. Podrobněji jsme o centrálních a distribuovaných modelech integrace hovořili v kapitolách 2.1.4.2 a 2.1.4.3.

Nejpoužívanějšími protokoly pro integraci informačních zdrojů jsou v současné době Z39.50 a OAI-PMH. Každý z nich zajišťuje jiné funkce, s oblibou se však v architektuře IS kombinují.

**Z39.50** slouží primárně pro vyhledávání na vzdálených serverech.<sup>52</sup> Data zůstávají na dílčích serverech, takže je integrační systém musí využívat tak, jak jsou k dispozici, a nemůže je nijak optimalizovat. Na vzdáleném serveru probíhá i vlastní vyhledávání – jedná se tedy i distribuované vyhledávání, kde centrální prvek pouze řídí a distribuuje vyhledávací proces a následně zpracovává vrácené výsledky. Pomocí Z39.50 můžeme budovat virtuální souborné katalogy a distribuované integrační systémy. V těchto systémech je vhodné nasadit metavyhledávače umožňující paralelní vyhledávání ve více zdrojích jediným dotazem.

Pozn.: Současný vývoj, který inklinuje k webovým technologiím (web services a XML), přináší nový standard SRW/SRU, který by měl nahradit starší a složitější Z39.50. V praxi se však standard SRW/SRU zatím příliš nerozšířil.

**OAI-PMH** neslouží k vyhledávání, ale primárně k hromadnému stahování (sklizení) metadat (popisných údajů informačních objektů) ze vzdálených serverů na lokální systém.<sup>57</sup> Filosofie vyhledávání je u OAI-PMH tedy opačná než u Z39.50: metadata jsou stažena na lokální systém a vyhledávání poté probíhá již lokálně nad vybudovanými lokálními indexy. Pomocí OAI-PMH můžeme budovat fyzické souborné katalogy a integrační systémy s centrálním úložištěm dat. Data (zejm. selekční údaje) lze optimalizovat a harmonizovat, případně obohacovat o další údaje, protože jsou fyzicky zcela pod kontrolou centrálního prvku. Vyhledávání tak může být oproti distribuovanému vyhledávání mnohem přesnější.

Více informací o standardech Z39.50, SRU/SRW a OAI-PMH přináší kapitola 3.2.

### 2.7.1.2 Vyhledávání podle zpracování vyhledávacích procesů

Významným technologickým hlediskem je také způsob, jak jsou dílčí informační zdroje prohledávány. Můžeme rozlišit v podstatě čtyři základní způsoby distribuovaného vyhledávání:

- **samostatně** (každý zdroj uživatel prohledává zvlášť, byť i ve stejném prostředí)
- **sekvenčně** (IS vyhledává ve vybraných zdrojích postupně)

- **pseudoparalelně** (IS vyhledává ve vybraných zdrojích střídavě a postupně tak zpracovává dotaz a výsledky; v jednu chvíli však komunikuje vždy pouze s jedním zdrojem; z hlediska uživatele se vyhledávání jeví jako paralelní)
- **paralelně** (IS vyhledává ve vybraných zdrojích současně)

Samostatné vyhledávání může být přesnější, protože umožňuje specificky definovat dotaz pro každý zdroj zvlášť s ohledem na specifický dotazovací jazyk daného zdroje. Zároveň však značně prodlužuje vyhledávací proces, zvlášť u velkého počtu požadovaných zdrojů, nehledě na to, že výsledky ze samostatných vyhledávání se pak těžko sjednocují a dále zpracovávají. Výjimkou je samostatné prohledávání centrálních úložišť nebo centrálních indexů, o kterých jsme se zmiňovali v kapitole 2.7.1.1. V integrovaném prostředí umožňující paralelní distribuované vyhledávání je vhodné zachovat také možnost samostatného vyhledávání ve zdrojích, aby mohl uživatel zadávat vysoce specifické dotazy, případně aby mohl použít nativní vyhledávací masku, tedy přejít do nativního rozhraní daného zdroje.

Paralelní vyhledávání umožňuje dynamicky seskupovat informační zdroje podle zájmu uživatelů, na místo statické nabídky podle dodavatelů a IS. Informační zdroje lze při paralelním vyhledávání libovolně kombinovat nezávisle na IS, vytvářet dynamicky oblasti zájmu podle potřeby.

**Paralelní vyhledávání** umožňuje zpracování dotazu, který směřuje současně na dva nebo více informačních zdrojů, a vrácený výsledek v reálném čase agreguje do jediné prezentace na výstupu. Může vzájemně kombinovat více protokolů, jako SRW, XQuery nebo Z39.50.

Paralelní vyhledávání se však potýká s řadou problémů, které zmiňujeme v dalších kapitolách. Jedná se zejm. o problém univerzálního vyhledávacího jazyka, dotazovacího jazyka, konverzí znaků a formátů, problém vícejazyčnosti dokumentů a následné zpracování a sjednocování výsledků vyhledávání.

### 2.7.1.3 Vyhledávání podle míry strukturovanosti informací

Vyhledávací techniky musí při integraci brát ohled na to, nakolik jsou informace v dílčích informačních zdrojích strukturovány. Informační zdroje mohou obsahovat:

- **nestrukturované** informace (pouze plné texty či jiné digitální objekty)
- **kombinace** strukturovaných a nestrukturovaných informací (plné texty či jiné digitální objekty doplněné popisnými metadaty – ty mohou být zapsány přímo v daném formátu)
- **strukturované** informace

Struktura informací napomáhá v sémanticky citlivém indexování a vyhledávání. Nutno dodat, že plné texty jsou dnes již zřídka kdy skutečně zcela nestrukturované. Např. HTML dokumenty obsahují značkování, kterým lze odlišit nadpis a tělo dokumentu, čehož lze využít při indexování a vyhledávání. Náznak sémantického značkování přináší různé popisná schémata postavená na XML, např. Dublin Core. Při integraci informačních zdrojů postavených na XML je však problém, že XML je standard zcela otevřený a dokonce i konkrétní schémata mají značnou volnost v interpretaci. O sémantickou interoperabilitu se snaží např. standard pro popis zdrojů RDF doplněný o prostředek RDFS pro definici RDF schémat.

**Klasický fond** lze automatizovaně prohledávat pouze za pomoci sekundárních informací (popisných údajů), které mají v knihovnách kvalitní a jemnou strukturu (dané zejm. knihovními standardy MARC, MDT, Konspekt apod.). Vyhledávání je proto značně přesné a sémanticky citlivé.

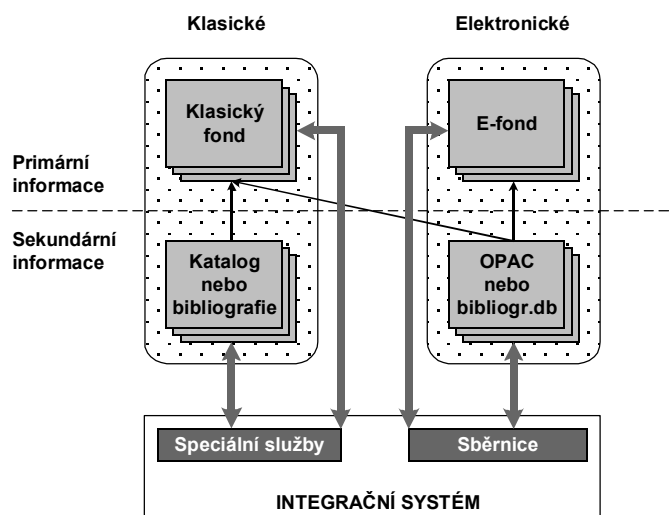
**Elektronické databáze a archivy** jsou ohraničeny v rámci konkrétního systému a indexovací stroj proto nemusí zkoumat a procházet nekonečný prostor dokumentů - předem ví, jaký obsah má indexovat. Indexování může probíhat jak v rámci plného textu, tak i s pomocí explicitních popisných údajů, kterými jsou většinou objekty v elektronických databázích a archivech opatřeny. Vyhledávací možnosti jsou proto nejširší ze všech tří množin zdrojů.

**Volný prostor internetu** je natolik rozsáhlý, že je základním úkolem vyhledávacích strojů vůbec jednotlivé dokumenty najít, aby je mohly zaevidovat a zaindexovat. Indexování probíhá automaticky na plném textu dokumentů, práce s metadaty a strukturovanými údaji je nedostatečná, pokud vůbec nějaká. Vyhledávání na webu je tak spíše vyhledáváním místa, kde se informace nacházejí, než hledání vlastních informací.

Snahou tvůrců standardů je zajistit u všech vznikajících informačních objektů použití jasných jednoznačných identifikátorů objektu. Na tomto poli již vznikla řada úspěšně nasazených a používaných evidenčních systémů – (e)ISBN, (e)ISSN, SICI atd., které se vztahují na konkrétní typy dokumentů. Postupně se zavádějí i univerzální jednoznačné identifikátory obecně pro jakékoli digitální objekty DOI, URN a PURL, či zastřešující standardy typu RDF. Zvláště ve volném prostoru internetu však vzniká obrovské množství dokumentů a informačních zdrojů zcela živelně a často neoficiálně. Jejich tvůrci své dokumenty nijak neoznačují ani nikde neevidují, takže jejich zapojení do vyhledávacího procesu je závislé na globálních vyhledávacích systémech typu Google, kde je úspěchem vůbec zjistit, že dané dokumenty existují.

#### 2.7.1.4 Vyhledávání podle míry digitalizace

Dalším významným hlediskem při vyhledávání je **míra digitalizace** informačních zdrojů. Je to bohužel jedním z nejpálčivějších problémů současných snah o integraci informačních zdrojů, protože zdaleka ne všechny informační zdroje jsou k dispozici v elektronické podobě. Vzhledem k jejich informační hodnotě však musíme s řadou z nich při integraci počítat. Připomeňme si schéma z kapitoly 2.1.3, kde jsme hovořili o procesní architektuře:

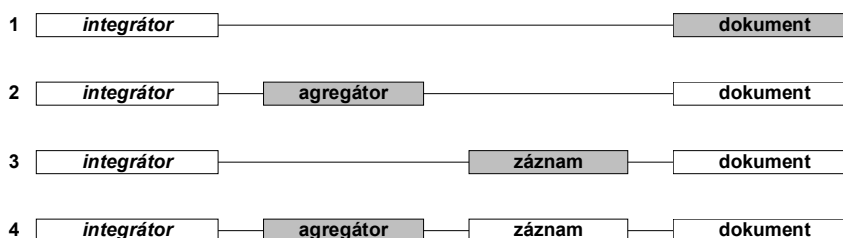


Primární i sekundární informace mohou být jak v klasické i elektronické podobě. Zatímco elektronické formy může integrační systém automatizovaně využívat, klasické analogové formy musí v integračním systému zprostředkovávat manuální obsluhu, čímž vznikají zejm. časové prodlevy a chybovost. Z hlediska integrace vnímáme tento stav jako slabinu a jsou proto silné snahy využívané a cenné klasické informační zdroje digitalizovat do elektronické podoby. Nicméně proces digitalizace je zdoluhavý a finančně náročný, proto existují určité mezistavy digitalizace jako přechodná řešení, např. oskenované katalogizační lístky nebo dokumenty bez převodu do textové podoby.

#### 2.7.1.5 Vyhledávání podle vztahu k primárním informacím

Jedním ze základních problémů integrace je, že různé informační zdroje obsahují různé typy informací s různým **vztahem k primárnímu informačnímu objektu** (dokumentu nebo obecně primární informaci): prohledáváme archivy plných textů, což jsou samy primární dokumenty, prohledáváme bibliografické databáze a katalogy, které obsahují pouze popis primárních dokumentů, prohledáváme integrované databáze, kde musíme nejprve nalézt dílčí zdroj, vyhledat záznamy a teprve následně vlastní dokumenty, prohledáváme soubory elektronických časopisů, kde můžeme hledat tituly časopisů ale současně v nich obsažené články apod. V zásadě můžeme popsat čtyři základní vztahy vyhledávacího stroje integračního systému

k prohledávaným objektům, které znázorňuje následující obrázek. Objekt „dokument“ zde pro zjednodušení reprezentuje nejen plnotextový dokument, ale jakýkoli primární informační objekt.



Prvním případem na obrázku je model, kde vyhledávací stroj prohledává **plnotextové dokumenty** nebo přímo prochází různé digitální objekty. To je případ vyhledávání v samostatných elektronických archivech bez účasti sekundárních informací, v dokumentových systémech či prohledávání volného prostoru internetu. Dokumenty mohou být strukturované i nestrukturované, při vyhledávání však nejsou používány žádné samostatné popisné (např. bibliografické) záznamy.

► Při vyhledávání nemusí uživatel hledat dílčí zdroje ale přímo hledá informace.

Druhý případ ukazuje model, ve kterém vyhledávací stroj prohledává prostřednictvím **agregátoru**, tedy určitého externího dílčího integrátoru, který zajišťuje vyhledávání v několika samostatných informačních zdrojích. Agregátory jsou typické zejm. pro úzce zaměřené oborové IS. Příkladem může být vyhledávání v článkové databázi, která v sobě integruje mnoho časopiseckých titulů, nebo integrovaný encyklopedický systém umožňující současné vyhledávání v několika dílčích online encyklopediích.

► Při vyhledávání nemusí uživatel hledat zdroje ale přímo informace ve všech zahrnutých zdrojích současně.

Třetí případ znázorňuje model, kde vyhledávací stroj vyhledává dokumenty prostřednictvím popisných (např. bibliografických) **záznamů**. Příkladem jsou OPAC, bibliografické databáze a nejrůznější evidence. Získání vlastních dokumentů musí následně zajistit navazující služby jako je linkování na primární dokument, EDD, výpůjčky atd. Tento způsob je také nejčastěji používaný pro prohledávání databází netextových digitálních objektů, které je nutné opatřit popisem, aby je bylo možné textově vyhledávat.

► Při vyhledávání musí uživatel nejprve najít relevantní dokumenty pomocí sekundárních informací, dokumenty poté musí získat přímo nebo formou služeb.

Čtvrtý případ znázorňuje model, kde vyhledávací stroj prohledává pomocí **agregátoru** popisné záznamy dokumentů. Agregátor opět vystupuje jako dílčí integrátor, jehož služby využíváme v našem integračním systému. Příkladem mohou být různé souborné katalogy nebo agregované bibliografie typu Medline.

► Při vyhledávání musí uživatel nejprve najít relevantní dokumenty pomocí agregovaných sekundárních informací, dokumenty musí následně získat přímo nebo formou služeb.

Samozřejmě existuje mnoho IS, které představují kombinace nebo mezistupně uvedených případů, např. databáze obsahující jak popisné záznamy, tak vlastní dokumenty, které proto nabízejí jak vyhledávání plnotextové, tak strukturované podle popisných polí, nebo např. terciální databáze, které shromažďují popisy jiných databází. Integrovat všechny tyto typy vyhledávání do jednoho systému do jednoho vyhledávacího modulu je nesmírně obtížné, ale zároveň klíčové, protože sjednocené vyhledávání představuje pravděpodobně nejprínosnější vlastnost IS pro integraci informačních zdrojů. Nejdůležitějším hlediskem je, jestli je **zdroj schopen vracet přímo primární informace nebo pouze jejich popis**.

Při integraci zdrojů vracejících primární a sekundární informace na jedné úrovni je důležité zajistit, aby uživatelé na výstupu jednoznačně rozpoznali, o jakou informaci se jedná a jak ji mohou dále využít. Kombinace primárních, sekundárních a případně terciálních informací je pro uživatele matoucí a nesystematické, nevrací se jim informace stejné kvality. Integrujeme-li zdroje v našem systému, vytváříme sami terciální informace (metadata) o zdrojích. Příkladem může být vyhledávač MetaLib použitý pro integraci informačních zdrojů v projektu JIB:

MetaLib (terciální) → OPAC (sekundární) → PDF (primární)

### 2.7.2 Dotazovací a selekční jazyky

O dotazovacích jazycích a uživatelském rozhraní, které s dotazovacími jazyky úzce souvisí, jsme se již zmínili v kapitolách 1.1.6 a 2.2. Při integraci informačních zdrojů se však problematika dotazovacích jazyků dále rozšiřuje, zvláště tehdy, když chceme používat jediné vyhledávací rozhraní pro vyhledávání ve více heterogenních informačních zdrojích.

Z tohoto hlediska musíme definovat, jaký je vztah **dotazovacího** jazyka k jazyku **selekčnímu**. Selekční jazyky slouží primárně k vyjadřování selekčních údajů<sup>32</sup>, tedy k popisu informačního objektu. Selekční údaje jsou následně využívány k vyhledávání. Dotazovací jazyky, které slouží ke komunikaci mezi uživatelem a IS, musí pro úspěšné vyhledávání pracovat se selekčními údaji daného informačního zdroje. Jinými slovy, dotazovací jazyky nabízejí způsob, jak se dotazovat (syntaxe), ale neřeší, na co se dotazovat (sémantika). Nevyjadřují, jak je konkrétní informační zdroj strukturovaný a jaký je jeho sémantický obsah. Každý dotaz vyjádřený prostředky konkrétního dotazovacího jazyka musí tedy obsahovat selekční údaje, které daný informační zdroj obsahuje, a musí respektovat jejich sémantiku. Příkladem může být IS využívající jako dotazovací jazyk SQL pro vyhledávání v databázi záznamů ve formátu UNIMARC, ve kterém jsou selekční termíny uloženy. Dalšími dotazovacími jazyky rozšířenými v oblasti knihovnictví a informačních zdrojů jsou Z39.50, CCL nebo moderní CQL používaný ve standardu SRU/SRW. Jako **identifikační** selekční jazyky jsou běžně používány různé varianty formátu MARC, DC nebo jiná schémata XML, METS, MIX (Z39.87), a jako **předmětové** selekční jazyky potom např. Konspekt, MDT, MeSH nebo různé systémy klíčových slov. Cílem této práce není hodnotit výhody a nevýhody jednotlivých selekčních jazyků, ani vyhledávacích technik. Při integraci informačních zdrojů nemůže ovlivnit, jakými selekčními systémy jsou dílčí zdroje vybaveny, musíme se snažit plně využít to, co nabízejí. Sladit různé selekční údaje pro potřeby paralelního vyhledávání je nelehký úkol a často si vyžádá tvrdé kompromisy.

Informační zdroje, která zapojujeme do integračního systému pro vzdálené vyhledávání, mohou být vybaveny selekčními systémy **prekoordinovaného** typu. Takový selekční systém však většinou nelze zabudovat do našeho IS a jsme nuceni jeho selekční údaje využívat postkoordinovaným způsobem, tedy hledáním podle uživatelem zadaných klíčových slov. To samozřejmě způsobuje vysokou nepřesnost, protože uživatel se málokdy „trefí“ do správné podoby prekoordinovaného výrazu. Tento problém může řešit centralizovaný způsob zapojení informačního zdroje, kdy budeme pravidelně sklízet jeho metadata a z nich potom budovat indexy. Tím můžeme vytvořit kompletní slovník selekčních údajů a nabízet v nich vyhledávání prekoordinovaným způsobem, např. použitím techniky procházení hesel (browsing, scan) či našeptávání (zobrazování nejbližších výrazů). Pokud není jasná struktura sklizených metadat, můžeme indexovat i plnotextově. Indexování je v tomto případě možné provádět pouze automaticky, zde se nám nabízí mnoho technik **automatické extrakce** pro výběr termínů vyjadřující obsah (lexikální analýzy, odstraňování nevýznamných slov, lematizace, srovnávání slov, frekvenční analýzy apod.) a **automatické přiřazování** termínu z řízeného slovníku na základě automatické obsahové analýzy.

V současné době, vlivem masivního rozšíření indexovacích technik internetových vyhledávačů, dochází k odklonu od tradičních prekoordinovaných selekčních systémů. Oblibu naopak získávají selekční systémy, jejichž nasazení lze plně automatizovat. Jsou to zejm. selekční jazyky pracující se slovy z názvů či jiných polí/atributů, pokud jsou ve struktuře informačního objektu k dispozici. Atomizace textu na slova s ohledem na strukturu (která často naznačuje sémantický význam) a následná indexace je dnes nejvyužívanějším způsobem automatického zpracování. Tyto systémy se také nejlépe integrují do jiných IS. Čistě postkoordinovaný přístup však vede k prudkému zvýšení selekčního šumu. Ten však může být potlačen použitím vhodného gramatického aparátu, nasazením různých technik rankingu, podporou jazykové flexe, pokročilými technikami vyhledávání apod.

Z hlediska uživatele obecně platí, že dotazovací jazyk by neměl být složitý, zároveň však musí mít dostatečný vyjadřovací aparát. Při formulování dotazu může být užitečným pomocníkem vyhledávací maska jako součást uživatelského rozhraní, která může uživateli proces zápisu usnadnit a posloužit jako průvodce definicí dotazu. Jedním ze základních úkolů uživatelského rozhraní je odstínit aplikační procesy vyhledávání, zejm. dotazovací jazyk používaný interně systémem (např. SQL), a nabízet uživateli vyhledávání jednoduchým, netechnickým způsobem. Dotazovací jazyk nabízený uživateli je v takovém případě zcela odlišný od jazyka, který používá systém v aplikačních procesech.

### 2.7.3 Problematika vyhledávací masky a dalších rozhraní pro vstup dat

Vyhledávací maska přesně definuje a zároveň limituje možnosti, jak uživatel formuluje svůj dotaz v integračním systému. Vyhledávací maska je nedílnou součástí uživatelského rozhraní, o kterém jsme hovořili v kapitolách 1.1.5 a 2.2. Je zároveň součástí dotazovacího jazyka, protože v masce uživatelé zadávají a vybírají údaje, ze kterých se následně poskládá finální dotaz. Vyhledávací maska musí tedy s dotazovacím jazykem korespondovat.

V IS pro integraci informačních zdrojů je vždy třeba vyřešit, jakým způsobem bude uživatel **vybírat informační zdroje**, které chce prohledávat. To souvisí se způsobem organizace informačních zdrojů v systému. Oproti vyhledávacím maskám jediného informačního zdroje tak uživatel musí nejprve vybrat zdroje a pak nad nimi definovat dotaz. Přibývá zde tedy další krok při formulaci dotazu, který musí být na úrovni uživatelského rozhraní a vyhledávací masky řešen tak, aby se uživatel v pracovním postupu dobře orientoval a věděl v každé fázi, jak má pokračovat.

Vyhledávací maska by měla podporovat různé úrovně uživatelů, což je nejčastěji řešeno jako dílčí masky pro snadné a pokročilé vyhledávání. Podívejme se však na nejběžnější formy trochu podrobněji.

Obliba **jediného políčka** pro zadávání dotazů je dnes díky webovým vyhledávacím typu Google vysoká. Výraz zadaný v tomto políčku se prohledává následně nad celou databází či databázemi a nad všemi podporovanými poli. Jednoduchost použití sebou však přináší velkou nevýhodu a tou je vysoká nepřesnost vyjádření. Ta se projevuje příliš velkou množinou vrácených záznamů s nízkou relevancí. Vyhledávací masky s jediným políčkem jsou proto často vybaveny **doplňujícími filtry**, kterými lze množinu prohledávaných informačních objektů zúžit (např. jazyk dokumentů, datum vzniku, formát souborů). Dotazovací jazyk na pozadí těchto masek také často podporuje syntaktické **prostředky k dalšímu zpřesnění dotazů** (např. ruční zápis uvozovek pro hledání výrazu jako fráze nebo použití booleovských operátorů). Tím uživatel používá syntax dotazovacího jazyka, např. CCL nebo SQL. Vyhledávací maska s jediným políčkem je v IS k dispozici jako tzv. jednoduché nebo snadné vyhledávání.

Pro zadávání přesnějších a složitějších dotazů slouží vyhledávací maska obsahující **více políček s volbou polí** a logické, většinou booleovské operátory. Masky může nabízet i další syntaktické prostředky (pravo/levostranné rozšiřování, maskování, použití proximitních operátorů atd.). Uživatel zde již může vizuálně vyjadřovat určitou sémantiku alespoň na úrovni polí, ve kterých se bude výraz vyhledávat, a tyto údaje logicky kombinovat. Může tak již definovat poměrně složité dotazy, např. pro vyhledání článků z oboru biochemie z roku 2004 a 2005. I zde mohou být k dispozici doplňující filtry a jednotlivá políčka mohou podporovat ruční zápis upřesňující syntaxe dotazu. Takováto vyhledávací maska je v IS často k dispozici jako tzv. pokročilé vyhledávání.

Některé systémy také nabízejí rozšíření funkcionality vyhledávací masky o tzv. **upřesnění dotazu**, kdy lze na vrácené množině dat provést další vyhledávání. Uživatel má tak možnost provést nejprve obecnější dotaz a v následujících krocích výsledek dále upřesňovat.

Zajímavou alternativou pokročilého vyhledávání jsou různé **vizuální průvodce formulací dotazu**, kteří laickým uživatelům umožňují komfortním způsobem definovat poměrně složité a přesné dotazy. V jednotlivých krocích je tak uživatel požádán, aby vybral prohledávané zdroje, zadal dotaz podle různých hledisek, zpřesnil dotaz v případě příliš velké množiny výsledků apod.

Kromě vyhledávacích masek pro uživatele jsou IS často vybaveny i dalšími typy rozhraní pro zadávání dotazů. Jsou to rozhraní určená pro komunikaci mezi systémy a slouží jako základní prostředek při integraci informačních zdrojů, kde propojujeme IS s mnoha vzdálenými informačními zdroji. Mezi tyto typy rozhraní patří zejm. brány Z39.50 a brány na bázi HTTP/XML.

Vraťme se však k problematice tvorby vyhledávacích masek pro uživatele a to jejímu nejpalčivějšímu problému. Jak jsme již naznačili na začátku této kapitoly a jak budeme podrobně probírat v kapitolách o konverzích, vyhledávací maska v integračním systému slouží pro zadávání jediného dotazu, který bude následně konvertován a rozeslán jednotlivým heterogenním informačním zdrojům. Tyto zdroje však mohou obsahovat informace, které lze vyhledat pomocí naprosto odlišných vyhledávacích bodů nebo pomocí vyhledávacích bodů s různě míněným významem. Např. v bibliografických databázích časopisů není přítomno pole

autor, naopak oproti knihám obsahují pole ISSN nebo např. pole obsahující název časopisu je jiné u záznamu článku a jiné u záznamu časopisu atd. Vyhledávací maska v integračních systémech musí být natolik univerzální, aby obsahovala **vyhledávací body**, které budou alespoň pro většinu typů informačních objektů společné. Tím je zaručena možnost integrace vyhledávání, ale vede ke ztrátě přesnosti vyhledávání. Lepších výsledků lze dosáhnout s dynamickou vyhledávací maskou, u níž se nabídka vyhledávacích bodů (políček s volbou polí) mění podle vybraných informačních zdrojů.

#### 2.7.4 Jazyk dokumentu a problematika vícejazyčnosti při paralelním vyhledávání

Při integraci informačních zdrojů se často setkáváme s problémem, že informační objekty obsahují informace v různých národních jazycích. Ve sbírkách můžeme mít dokumenty v češtině, angličtině, ruštině, japonštině a dalších jazycích. Jaké jsou možnosti tyto dokumenty vyhledávat jediným dotazem? Obecně musíme odlišit, zda se jedná o plnotextové vyhledávání ve vlastním dokumentu nebo vyhledávání prostřednictvím sekundárních informací.

U **plnotextové** vyhledávání se musíme spolehnout na implementaci automatických překladů dotazu do cílových jazyků za pomoci překladových slovníků. Narážíme však na problémy automatického zpracování sémantiky přirozených jazyků, které nejsou doposud v praxi uspokojivě řešeny\*. U vědeckých prací je proto často k dispozici překlad anotace do některého světového jazyka, zejm. angličtiny, který zhotovil profesionální překladatel nebo odborník se znalostí daného jazyka, čímž je při vyhledávání zaručeno zachování smyslu a významu. Nejideálnějším případem je pak samozřejmě kvalitní překlad celého dokumentu.

U **sekundárních** informací, které dokument popisují, se nabízí více možností. Kromě nedokonalých automatických překladů, je možné dokumenty popisovat pomocí vícejazyčných tezaurů nebo jiných řízených systémů předmětových hesel či klíčových slov, které jsou předem kvalitně připraveny za pomoci odborníků s výbornou znalostí daných jazyků. Vyhledávací systém je pak schopen vyhledávat pomocí hesel ve všech jazycích, které vícejazyční popisný systém podporuje. Významným počinem v této oblasti je např. projekt Vícejazyčný předmětový přístup do knihovních katalogů – MSAC (Multilingual Subject Access to Library Catalogues) a mnoho dalších. Významnou pomoc zde nabízejí tradiční klasifikační systémy, jako je MDT, kde je nositelem významu jazykově abstraktní kód, takže je možné je využívat v kterémkoli národním systému.

Samostatným problémem, který souvisí s národními jazyky, jsou různé systémy písma (latinka, azbuka, čínské písmo), směr zápisu písma (zleva doprava, zprava doleva) a kódování znaků.

#### 2.7.5 Konverze znakových sad

Jak bylo naznačeno v kapitole 2.7.4, při integraci informačních zdrojů se můžeme setkat s dílčími zdroji dat v rozmanitých národních jazycích a znakových sadách (jen pro češtinu existuje 6 znakových sad!). Při distribuci a transformaci dotazu je často třeba konvertovat znaková sady, aby při komunikaci dotazu mezi uživatelem a cílových vyhledávacím strojem nedošlo k nežádoucí deformaci znaků. Tento jev je poplatný znakovým sadám využívající pouze 256 znaků. Pokud např. uživatelský klient používá kódování windows-1250 a cílová databáze iso-8859-2, bez správné znakové konverze by systém neprovedl vyhledání správně. Při použití národních 256-znakových sad však nelze korektně převádět znaky mezi jednotlivými rodinami jazyků (např. mezi východoevropskými a západoevropskými), protože žádná národní sada neobsahuje znaky všech národních abeced.

Tyto problémy úspěšně řeší univerzální znaková sada **Unicode** (resp. jeho nejrozšířenější modifikace UTF-8), která rozšířila počet znaků v sadě na 38 885 (Unicode verze 2.0). Tím jsou v této sadě pokryty prakticky všechny znaky národních abeced světa a různé speciální znaky a symboly. Potřeba konverzí tím zcela odpadá. Stále více systémů a informačních zdrojů prezentuje svá data v UTF-8.

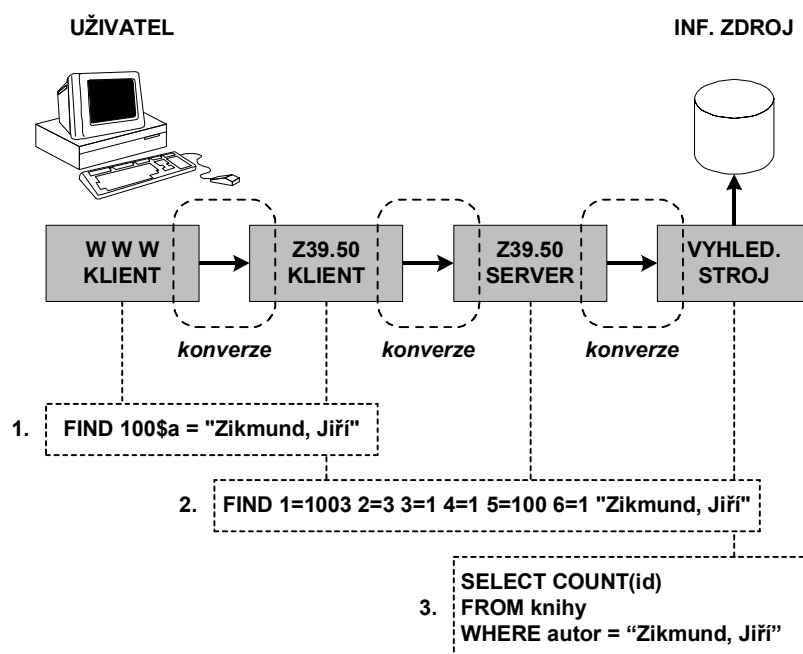
\* Objevují se však zajímavé experimentální projekty, které se snaží o vývoj vyhledávacích strojů schopných prohledávat vícejazyčné dokumenty pomocí sémantické analýzy a následného překladu do více jazyků, např. projekt M-CAST (Multilingual Content Aggregation System) využívající vícejazyčný sémantický vyhledávací stroj TRUST.

Jeden problém však nadále přetrvává, a tím je **způsob zápisu** různých národních znaků na klávesnici. V sadě Unicode 2.0 sice máme 38 885 znaků, ale těžko si představit, že by měl uživatel na stole klávesnici s 38 885 klávesami! V operačních systémech je možné podporu jednotlivých národních jazyků přepínat, čímž se nám aktivují příslušné podskupiny znaků ze sady Unicode, ale tím se mění význam kláves a bohužel i rozložení, protože každý národ většinou používá rozložení znaků na klávesnici podle staré konvence odvozené od mechanických psacích strojů. Popisky kláves mohou být max. dvojjazyčné. Jako pomůcka se proto často používá tzv. virtuální klávesnice, která se uživateli zobrazí na obrazovce. V ní může buď přímo myší označovat klávesy a tímto způsobem psát nebo slouží jako nápověda k rozložení znaků a uživatel podle ní píše na skutečné klávesnici. Jiným řešením může být podpora vyhledávání bez diakritiky, alespoň v rámci stejných systémů písma (např. latinky), je tím však snížena relevance a přesnost vyhledávání.

### 2.7.6 Konverze a transformace dotazů s ohledem na selekční údaje a indexy

Zadá-li uživatel v integračním systému dotaz, musí integrační systém pro každý vybraný informační zdroj **zkonvertovat** dotaz do dotazovacího jazyka s podporou selekčních údajů, které daný zdroj používá. Nestačí konvertovat dotazy a protokoly, ale je třeba brát ohled na používané databázové indexy. Nejlépe si vše ukážeme na příkladu:

Čtenář chce vyhledat knihy od spisovatele Oty Pavla. Do vyhledávací masky integračního systému zapíše do pole *Autor* výraz „Ota Pavel“. Uvozovky zde označují, že se má výraz hledat jako přesná fráze – bez uzavření do uvozovek by se hledala izolovaná slova a nesprávně by se vyhledal např. dokument od dvojice autorů Ota Vrbský a Pavel Boroš. Čtenář pro vyhledávání vybere 2 informační zdroje – katalogy dvou různých knihoven (jeden katalog používá MARC 21, zatímco druhý katalog UNIMARC). Integrační systém musí na pozadí údaje z vyhledávací masky zkonvertovat do své vnitřní syntaxe, kterou může být např. jazyk CCL. Aplikace pak musí CCL překonvertovat do formátu Z39.50, kterým IS komunikuje s oběma vzdálenými katalogy. Nestačí však pouze konvertovat dotazovací jazyk, musí upravit také obsah, protože katalogy sémanticky odlišně naplňují atributy Z39.50. Konkrétně jeden katalog v atributu 1003 (autor) nabízí výraz *Pavel, Ota* (přesně jak je zapsáno v podpoli 100\$a v MARC 21) a druhý katalog výraz *Ota Pavel* (poskládáním z podpolí 700\$a a 700\$b v UNIMARC). K dalším problematickým polím u bibliografických databází konvertovaných do Z39.50 patří tradičně předmětové heslo a klíčová slova (každá databáze může být jiný systém předmětových hesel nebo jiná pravidla pro tvorbu klíčových slov), zápis roku/rozpětí let a ISBN/ISSN (indexování s pomlčkami, mezerou nebo bez mezery).



Obr. 14 - Příklad transformace dotazu



Z příkladu můžeme odvodit dvě významné skutečnosti. Při integraci informačních zdrojů musíme při komunikaci řešit nejen **konverze různých dotazovacích jazyků**, ale i **konverze vlastních selekčních údajů**. Nemůžeme se spoléhat pouze na platné standardy, protože žádný standard pro svoji obecnost neřeší podrobně uvažovaný obsah, tedy sémantiku. Když problematiku dotazovacích jazyků ještě rozšíříme o problém **vícejazyčnosti obsahu a nejednotnosti věcného i jmenného popisu** u záznamů a metadat informačních objektů, získáváme poměrně komplikovaný balík problémů, které musí systémoví architekti řešit.

**Sémantiku** jednotlivých metadatových polí řeší různé selekční systémy a na ně aplikovaná katalogizační pravidla, pro formát MARC jsou to např. AACR2, nicméně při transformacích se často ztrácí jejich účinnost. Problém se projevuje nejvíce u polí věcného popisu, kdy se např. do atributu „subject“ (21) standardu Z39.50 mapují v různých IS zcela odlišné selekční systémy: různé kvalifikační systémy (MDT), předmětová hesla, výrazy tezauru, klíčová slova apod. Jejich zápis může být navíc rozmanitý vlivem transkripce, použití národních jazyků či místních zvyklostí. Stejný problém nastává u DC a dalších méně strukturovaných formátů. Při využití informačních zdrojů s tak velkými odlišnostmi při transformacích dotazů se při paralelním vyhledávání vracejí zcela nesourodé a nesrovnatelné množiny záznamů. Relevance vyhledávání je ve výsledku velice nízká. U každého vyhledávacího bodu, který chceme při paralelním vyhledávání použít, je proto důležité nejen jestli v daném informačním zdroji vůbec **existuje**, ale i jaký je jeho **obsah**.

Vyhledávání však málokdy probíhá přímo nad vlastním datovým úložištěm. K vyhledávání se vytvářejí vyhledávací indexy, které jsou podle definovaného nastavení budovány z uložených dat. Např. z uloženého bibliografického záznamu se vytváří indexy autorů, názvů apod., nad kterými pak probíhá vlastní vyhledávání. Neexistuje však žádný obecně platný standard, který by popisoval, jak mají být indexy budovány. Různá řešení tak mohou indexovat tentýž obsah zcela odlišným způsobem, čímž jsou v závěru zmarněny snahy o unifikaci polí a obsahu ve vlastních datech. Při spolupráci členů zapojených do integračního systému je proto nanejvýš vhodné definovat **pravidla pro sjednocení indexů a vyhledávacích bodů**. Mezi taková pravidla může patřit použití malých a velkých písmen při indexování, nasazení řízených slovníků/autorit s definicí jejich indexování, ošetření speciálních znaků, způsob zpracování slov a frází, interpretace operátorů, proximita, podpora jazykové flexe apod. Pro správnou funkci vyhledávání v IS musí být tedy dokonale sladěna struktura dat s jejich obsahem, se způsobem indexování těchto dat ve vyhledávacím systému a vlastním dotazovacím jazykem.

Indexy buď můžeme využívat vzdáleně (např. Z39.50) anebo lokálně (např. indexy sklízené přes OAI-PMH). **Vzdálené indexy** mají sice tu výhodu, že se o jejich budování a správu nemusíme starat, jsou pro nás již připraveny, ale nemáme většinou vliv na jejich strukturu ani obsah. **Lokálně spravované indexy** (ať již indexy vlastních lokálních zdrojů nebo indexy stažené ze vzdálených zdrojů) máme naopak zcela pod kontrolou a můžeme je libovolně upravovat a optimalizovat.

Indexování může narážet i na technické problémy digitálních objektů, které indexaci prakticky znemožní. Typickým příkladem je dokument HTML, kde jsou nadpisy z důvodu designu řešeny jako grafické bitmapové objekty. Podobný problém se objevuje u dokumentů PDF, kde může být celý text převeden na křivky. HTML naštěstí nabízí nástroj, jak grafické prvky vyjádřit zároveň textově, a to použitím tagů ALT či TITLE – indexovací stroj může tyto tagy použít namísto grafické prezentace. Pokud integrační systém musí zpracovávat texty prezentované pouze graficky, zbývá už pouze možnost rozpoznávání znaků OCR nebo ruční indexace.

### 2.7.7 Zpracování a prezentace výsledků

Při prohledávání různých typů informačních zdrojů získáváme na výstupu odlišné typy informací. Uživatel je z této nesourodé množiny výsledků často natolik **zmaten**, že neví, jak má dále pokračovat v procesu získání primárního dokumentu. Integrační systém, zvláště v případě, že podporuje paralelní vyhledávání, musí uživateli na výstupu nabídnout konsolidované výsledky v takové formě, aby se uživatel okamžitě zorientoval.

Tvorbu prezentace výsledků vyhledávání můžeme rozdělit na 2 fáze: na **fázi zpracování výsledků**, kterou zajišťuje systém na pozadí na základě vrácených výsledků vyhledávání, a na **fázi generování zobrazení výsledků**, které vidí uživatel na výstupním zařízení.

### 2.7.7.1 Zpracování výsledků vyhledávání

Vrácené výsledky vyhledávání z jednoho nebo více informačních zdrojů by měl integrační systém před vlastním zobrazením uživateli zpracovat a unifikovat, aby uživatel při každém vyhledávání získal srovnatelné výsledky. Zpracování vrácených záznamů může obsahovat mimo jiné následující dílčí operace:

- **deduplikace** - odstraňování více výskytů totožného informačního objektu
- **slučování** – prezentace podobných informačních objektů jako jeden objekt při zachování jejich odlišných vlastností (např. jeden záznam s uvedením více sigel vlastníků)
- **optimalizace** – data jsou upravována tak, aby byla pro uživatele lépe čitelná (např. sigly knihoven jsou nahrazeny oficiálními názvy knihoven)
- **harmonizace** – informační objekty jsou převáděny do formy, která odpovídá standardům a zvyklostem daného integračního systému. Harmonizace může probíhat na úrovni podpolí, atributů nebo celých záznamů. Příkladem může být převod jména do inverzního tvaru, převod do struktury odpovídající MARC 21 apod.
- **obohacování** – informační objekt je uměle obohacen o informace z dalších informačních zdrojů, takže vzniká objekt zcela nové kvality (např. k bibliografickým záznamům jsou připojeny recenze z databáze recenzí)

Takto zpracované výsledky jsou připraveny pro fázi zobrazování. Proces zpracování výsledků vyhledávání probíhá **v reálném čase** mezi okamžikem vrácení výsledků z informačních zdrojů a okamžikem zobrazení. Některé takové operace lze však provádět již před vlastním vyhledáváním a to v případě, že jsou **data předem lokálně stažena** (např. při zpracování sklizených dat přes OAI-PMH). Při zpracování v reálném čase je samozřejmě důraz na minimalizaci časových prodlev, takže často není možné provést tolik operací nad vrácenými záznamy, jako je to možné u předem stažených nebo lokálních dat.

### 2.7.7.2 Zobrazování výsledků vyhledávání

Proces zobrazování výsledků uživateli, tedy vlastní prezentace dat, je determinován uživatelským rozhraním a jazykem tohoto rozhraní (více informací o uživatelském rozhraní – viz kapitola 2.2). Při zobrazování je klíčové, jakým způsobem budou data rozčleněna a uspořádána a jakým způsobem bude moci uživatel na vrácených datech provádět další operace. V orientaci mohou pomoci nápovědy a různé vysvětlovací komentáře.

**Formát zobrazení** rozhoduje o tom, jakým způsobem budou zobrazeny vlastnosti informačního objektu. Může to být tabulkové zobrazení, kde jsou jednotlivá pole záznamů zobrazena ve sloupcích, může to být citační zobrazení, kde jsou všechna pole zobrazena v citační větě, mohou to být různé vizualizace apod. Uživatel by měl mít možnost formáty zobrazení měnit podle potřeby a jako výchozí by mělo být nastaven takový formát, s jakým je spokojena většina uživatelů IS. Formát zobrazení by měl odpovídat platným normám a standardům, pokud jsou pro daný formát k dispozici (např. bibliografické citace).

**Řazení** určuje, jakým způsobem budou seřazeny vyhledané informační objekty ve výsledné množině za sebou. Uživatel by měl mít možnost měnit hlediska řazení, aby mohl vyhledané výsledky pohodlně a dynamicky přeřazovat. Typicky se výsledky řadí abecedně podle jména autora či názvu dokumentu, numericky podle roku vydání apod. K řazení lze však využít i jiná hlediska než jsou pole či vlastnosti nalezených objektů. Může to být např. míra relevance nalezených objektů (vypočítaná na základě zadaných algoritmů při zpracování výsledků) nebo podle možnosti získat primární dokumenty, aby se uživatelům na prvních pozicích seznamu nabízely výsledky s nejsnadnější a nejrychlejší cestou k plnému textu či digitálnímu objektu.

**Třídění** je proces, který rozděluje množinu výsledků vyhledávání do podmnožin podle určitých hledisek a tyto podmnožiny potom přehledně zobrazuje. Typicky mohou být záznamy v podmnožinách seskupeny podle typů dokumentu, jazyka, podle roku vydání, podle autorů atd. Třídění podle konkrétního pole nebo vlastnosti informačního objektu vytváří tzv. fasety. Třídění, které pracuje na základě sémantické analýzy vrácených informačních objektů, se nazývá klastrování a začleňuje vyhledané záznamy do vygenerovaných skupin předmětových hesel podle informací, které v sobě obsahují. Klastry a fasety mohou sloužit i jako filtry (viz filtrování).

Dotaz: předmět = biometrika Celkem nalezeno: 345 záznamů	Biometrika (345) - teorie (112) - biometrika a bezpečnost (65) - otisky prstů (16) - identifikace podle oční rohovky (47) - biometrické snímače (29) - ostatní (76)
---	---

Obr. 15 - Rozložení výsledků do předmětových podmnožin

**Filtrování** pomáhá uživateli lépe se orientovat ve výsledcích tím, že zobrazuje pouze ty vyhledané objekty, které odpovídají použitému filtru. Filtry lze navíc různě kombinovat. Můžeme např. z vyhledaných záznamů dokumentů zobrazit pouze ty, které jsou psány česky a vyšly v roce 2006. Filtry mají výhodu, že se aplikují na množině již vyhledaných záznamů, aniž bychom museli spouštět další vyhledání. Jako dynamická nabídka filtrů mohou dobře posloužit již zmíněné fasety a klastry.

Ve fázi prezentace musí být přehledně nabízeny navazující služby, viz kapitola 2.8.

Dotaz: předmět = biometrika	
Tomuto dotazu vyhovuje:	27 dokumentů 98 záznamů o dokumentech 11 dalších informačních zdrojů, kde lze získat další informace
Chcete:	zobrazit dokumenty? zobrazit záznamy s možností vyhledání lokace dokumentů? prohledat další informační zdroje?

Obr. 16 - Rozložení výsledků podle typu informací s vazbou na navazující služby

Jinou možností je už na úrovni zadávání dotazů striktně rozdělit skupiny zdrojů podle typu informací a umožnit paralelní vyhledávání pouze u zdrojů stejného typu. Tj. samostatně prohledat archivy plných textů, samostatně zdroje bibliografických informací a zvláště terciální databáze. K tomu samozřejmě mohou přibýt skupiny speciálních databází, např. databáze osobností či různé ekonomické rejstříky. Zobrazení výsledků je pak srozumitelnější, ale uživatel musí případně prohledávat mnoho typů dokumentů extra. Tato volba by měla být nabízena pouze jako jedna z možností.

### 2.7.7.3 Vyhledávací pomůcky

Vyhledávací proces je často proces složitý a uživatel se nedobere k požadovaným výsledkům na první dotaz. Uživatel si během vyhledávání většinou teprve formuluje a upřesňuje svoje požadavky. Proces vyhledávání by nikdy neměl skončit ve slepé uličce s konstatováním, že nebyly nalezeny žádné relevantní záznamy. Velmi užitečné je proto nabízet spolu s výsledky vyhledávání i různé pomůcky pro další postup. Mluvíme o vyhledávacích pomůckách, které představují speciální typy integrovaných služeb.

Základní vyhledávací pomůckou jsou **návrhy úpravy dotazu**. Ty mohou pracovat na základě rozpoznání překlepů, např. nabídkou podobných výrazů ze slovníku (při zadání výrazu „beethoven“ pomůcka nabízí změnu výrazu na „beethoven“), nebo na základě analýzy historie dotazů ostatních uživatelů (pomůcka napovídá: uživatelé, kteří jako vy hledali „voskovec“, také hledali „werich“).

Jako **základ pro další vyhledávání** mohou také sloužit samotná vyhledaná data. Některá pole vyhledaných záznamů mohou např. pracovat jako aktivní odkazy spouštějící vyhledání podle obsahu vybraného pole. Vznikají tak hypertextové vazby mezi informačními objekty.

Při neúspěšném hledání mohou být k dispozici **scénáře dalšího postupu**, případně přesměrování na referenční služby, kde mohou uživatelé volně formulovat své požadavky a pracovníci knihovny jim v co nejkratší době pomohou s řešením.

## 2.8 PROPOJENÍ INFORMAČNÍCH ZDROJŮ A SLUŽEB

Již jsme zmínili, že hlavním cílem při využívání informačních zdrojů je dodat uživateli primární informace. Také jsme si ukázali, že ne každý informační zdroj dokáže dodat primární informace přímo. IS pro integraci informačních zdrojů by však měl být schopen nabízet uživateli navazující služby, které ho k primárním informacím přivedou a dokončí tak celý proces uspokojování informačních potřeb. Paleta služeb může být samozřejmě daleko pestřejší - může např. přinášet řadu nástrojů pro vyhledávání příbuzných dokumentů pomocí bibliografických a citačních vazeb, prostředí uživatelské schránky, mikroplatební systémy pro realizaci dodávání dokumentů za úhradu apod.

Nabízené služby nemusí být nutně realizovány přímo v našem IS. Naopak je často velice výhodné **využívat služeb externích IS**, buď zdarma nebo za úplatu, protože často dosahují takové kvality, které bychom sami z různých důvodů těžko dosahovali. Jak lokální, tak i externí služby by však měly být **bezešvě integrovány** na našeho IS, aby uživatel mohl se všemi nástroji pracovat na jednom místě a v jednom prostředí. Důležitými momenty je proto integrovaná autentikace (SSO) a zapojení všech komponent do jednoho uživatelského prostředí, k čemuž nejlépe poslouží portálové technologie (více informací o portálech – viz kapitola 2.10).

Informační služby, které mají být využitelné v jiných IS, musí být k tomuto účelu vybaveny vhodným systémovým rozhraním. Ke vzájemné spolupráci a integraci více IS je nutné, aby rozhraní jejich služeb byla univerzální a aby jim ostatní IS porozuměly. Proto začaly vznikat snahy o standardizaci služeb a jejich popisu. Nejrozšířenější technologií v této oblasti jsou dnes **webové služby** (podporované W3C) využívající většinou komunikační protokol SOAP a jazyk pro popis služeb WSDL (Web Service Description Language). SOAP i WSDL jsou postaveny na XML a důsledně využívají Unicode, čímž je zajištěna jejich univerzálnost a čitelnost. U jednodušších aplikací s webovým rozhraním se s oblibou místo poměrně složitého protokolu SOAP využívá protokol HTTP s posíláním parametrů přes URI a reprezentací v XML, HTML apod. Na tomto principu jsou založeny služby využívající protokoly REST nebo OpenURL. Popis služby ve WSDL definuje, jaké rozhraní služba nabízí, tj. jaké operace a jejich parametry očekává a jaké hodnoty vrací, a také specifikuje protokol a adresu, kde je služba dosažitelná. Aby IS popisy služeb našly, vznikají prohledatelné seznamy webových služeb. Po vyhledání je již snadné danou webovou službu použít a volat z libovolného IS. I proto je u webových služeb, které přenášejí důvěrné informace, nutné více než kde jinde důsledně dbát na bezpečnost a ochranu přenášených dat.<sup>25</sup>

Z hlediska implementace služeb v prostředí IS, obecně platí, že by měly být k dispozici vždy kontextově, tj. měly by být v nabídce právě a pouze v okamžiku, kdy jsou pro uživatele relevantní a užitečné. Např. nabídka služby elektronického dodávání dokumentů by se měla nabízet jen v případě, kdy k vyhledanému záznamu článku neexistuje přímý odkaz na plný text. Taková nabídka by se naopak neměla zobrazovat v případě, jedná-li se o zvukový nosič apod.

Následující kapitoly ukáží základní a nejpřínosnější typy služeb, které mohou výrazně zvýšit užitečnost a pracovní komfort IS pro integraci informačních zdrojů.

### 2.8.1 Služby postavené na vyhledaných objektech

Základním typem služeb, které můžeme nabízet v IS pro integraci informačních zdrojů, jsou služby, které využívají údajů z již nalezených informačních objektů. Takovými objekty mohou být např. bibliografické záznamy, různé adresářové záznamy apod. Údaje z objektů poslouží jako vstupní parametry služby, která pak uživateli nabídne další doplňující informace nebo ho nasměruje na další speciální služby, jako je např. EDD. Služby postavené na vyhledaných objektech nejčastěji využívají tzv. bibliografické či citační vazby nebo princip linkování.

### 2.8.1.1 Bibliografické vazby

Bibliografické vazby využívají vztahů v rámci polí či atributů jmenného nebo věcného popisu, tedy sekundárních informací. Nejčastěji se takto využívá polí původců děl, názvů děl a edic, předmětových a jiných věcných kategorií. Díky bibliografickým vazbám může uživatel obohatit proces vyhledávání o možnost dohledat příbuzné dokumenty z již vrácené množiny záznamů nebo se dostat k primárním informačním objektům, nejčastěji plným textům dokumentů. Bibliografické vazby mohou být v IS realizovány jako hyperlinky na polích záznamů v seznamu výsledků nebo pomocí linkovacího serveru.

**Linkovací server** zajišťuje typ služeb, které umožňují uživateli přejít od vyhledaných informací o informačním objektu přímo k danému objektu. Typický je to přechod od bibliografického záznamu přímo k plnému textu. Linkovací servery využívají nejčastěji protokol OpenURL, který slouží k přenosu bibliografických informací. Typickým příkladem nasazení je Google Scholar, ve kterém uživatel vyhledá údaje o akademických či vědeckých pracích a linkovací server následně vyhledá lokace jejich exemplářů. Stejným způsobem může linkovací server spolupracovat s katalogy knihoven či jinými bibliografickými nebo jinými databázemi. Alternativou linkovacího serveru jsou přímé odkazy na plné texty jako součást samotného popisného záznamu.

Generování bibliografických vazeb je zcela závislé na kvalitě popisu jednotlivých záznamů, která často úzce souvisí s problémy nevhodných katalogizačních pravidel, rozdílných pracovních postupů, subjektivitu a pečlivosti jednotlivých katalogizátorů či editorů obsahu. Přestože jsou pro popis zaváděny řízené slovníky, autority a tezaury usilující o jednoznačnost výrazů a entit, proces popisu je subjektivní (už jen proto, že editor často podrobně nezná popisované dílo).

Bibliografické vazby nejlépe fungují na polích, kde se používají jednoznačné identifikátory, např. ISBN/ISSN, či obsah řízený slovníkem, typicky autoritami. Když je jednoznačný obsah pole, jsou jednoznačné i vazby. Proto se zejm. v oblasti knihovnictví usiluje o zavádění standardních identifikátorů i pro jiné typy informačních objektů, než jsou pouze monografie a periodika. Očekává se např. schválení a zavedení ISADN (International Standard Authority Data Number) pro identifikaci všech autoritních údajů, standardizace a rozšíření SICI pro jednoznačnou identifikaci článků či obecného identifikátoru DOI pro jakýkoli digitální objekt. Iniciativ v této oblasti nalezneme mnoho, při implementaci je však vhodné použít pouze takové identifikátory, které jsou v praxi rozšířené a mají naději na přežití.

Klasické bibliografické vazby, vycházející většinou z definice použitého formátu (MARC, DC), jsou vertikálně ploché a nedokáží vyjadřovat vztahy mezi dílem, realizací díla, jeho fyzickým provedením a reprezentací v podobě konkrétní jednotky. Tento problém úspěšně řeší organizací IFLA definovaný bibliografický standard **FRBR** (Funkční požadavky na bibliografická záznam). FRBR rozeznává 3 skupiny entit: původce, jejich produkty a okolnosti vzniku. Ve skupině produktů rozeznává díla (jako abstraktní intelektuální či umělecký výtvar), exprese (vyjádření díla), manifestace (hmotné provedení) a samotné jednotky jako konkrétní informační objekt. Lze tak např. přehledně vyjádřit vztah mezi spisovatelem, jeho románem, překlady románu do různých jazyků, konkrétními knižními vydáními v různých jazycích a konkrétními knihovními jednotkami, které si může čtenář vypůjčit. Při vyhledávání, které podporuje dynamické vazby podle FRBR, čtenář vidí přehledný souhrn všech relevantních reprezentací daného objektu, např. knižní vydání ve všech jazycích (bez ohledu na jazyk použitý při vyhledávání nebo předmětová hesla). Rada IS již FRBR podporuje, např. AKS Virtua od VTLIS či Primo od Ex Libris.

Podobně jako bibliografické vazby mohou v nebibliografických systémech sloužit vazby mezi jakýmkoli atributy nebo údaji. Příkladem mohou být vazby mezi adresářovými položkami a GIS (na základě uvedené adresy firmy nám systém zobrazí polohu firmy na mapě), mezi geografickými údaji a dopravními informacemi apod.

### 2.8.1.2 Citační vazby

Citace představují přirozené vazby mezi dokumenty na základě intelektuálního využití celého nebo části díla při tvorbě díla nového. Citace je tedy údaj o dokumentu, který autor bezprostředně použil při přípravě své práce<sup>32</sup>. Na rozdíl od bibliografických vazeb, které vycházejí pouze z popisu díla (který je často hrubozrný a vytvářejí ho lidé, kteří popisované

dokumenty nečetli a neznají detailně problematiku), vytvářejí citace jasné obsahové vazby a navíc ukazují jasný směr vlivu (citovaný vs. citující). Citující zná dobře citované dílo a jeho využití často poukazuje na dobrou kvalitu citovaného díla. Citační a bibliografické vazby by se měly doplňovat.

Citační vazby bychom neměli zaměňovat s různými technologiemi sledování webových odkazů jako je např. Google PageRanking. Zatímco vědečtí pracovníci téměř vždy citují hodnotná a relevantní díla, takže citační vazby nás navádějí na další kvalitní práce, na webu se objevují odkazy, jejichž obsah často vůbec nesouvisí s původní stránkou. Tvůrci webu k tomu mají mnoho různých důvodů (placené odkazy, odkazy na „spřátelené“ weby atd.). Naopak na důležité relevantní stránky z různých důvodů nemusí odkazovat.

Citační vazby buď můžeme analyzovat přímo v integračním IS, pokud nám to povaha systému dovolí (např. v případě systému pro evidenci vysokoškolských kvalifikačních prací) nebo můžeme využívat a integrovat služby komerčních citačních systémů. Mezi nejvyužívanější citační databáze dnes patří Web of Science a Scopus.

### 2.8.2 Služby spojené s uživatelskou schránkou - personalizované služby

Nabídka uživatelské schránky přináší uživatelům možnost ukládat si přímo na serveru personalizované nastavení a výsledky své práce a plnit tak pomocí IS dlouhodobější úkoly. Předpokladem zřízení uživatelské schránky je registrace uživatele v IS. Využití uživatelské schránky může být velice široké, typickými aplikacemi jsou následující služby:

- Uživatelské citace (a jejich propojení na osobní citační a bibliografické manažery)
- Osobní bibliografické seznamy (citace, doporučená literatura atd.) s možností obohacování záznamů o osobní poznámky
- Uložené dotazy, SDI, osobní vyhledávací profily
- Datová úložiště (na způsob FTP)
- Osobní komunikace (e-mail, chat, diskuse)
- RSS a informační avíza
- Výstřížkové služby (NetSnippets, Clipmarks)
- Publikování (připojování recenzí na knihy, osobní stránky, sdílení obsahu)
- Speciální služby podle zaměření (kalendáře, plánování, náhledy do jiných IS, např. do intranetu)
- Přednastavení a kustomizace prostředí IS

Na základě registrace a následné rozpoznání identity uživatele může IS nabízet dedikované služby, které lépe odpovídají požadavkům a zvyklostem uživatele. Vše může být „ušito na míru“ konkrétního uživatele – uživatel není zatěžován informacemi a službami, o které nemá zájem, naopak dostává přehledně a efektivně vše, co potřebuje.

Předpokladem spolehlivé funkce uživatelských schránek je bezpečnost, respektování diskrétnosti jejich obsahu a spolehlivost. Zároveň je třeba zajistit (nejlépe jako podmínka registrace), aby uživatelé souhlasili s dodržováním provozních podmínek, tj. aby jejich aktivity spojené s uživatelskou schránkou nebyly v rozporu s dobrými mravy, platnými zákony a obecně závaznými etickými a morálními principy společnosti nebo nesloužily k páčání trestné činnosti.

### 2.8.3 Služby spojené s dodáním primárních informací

Služby dodávání primárních informací patří mezi nejzákladnější a uživatelsky nejžádanější typy služeb, protože přinášejí uživatelům to, co nejvíce vyžadují – primární dokument, digitální objekt či faktografickou informaci. Bohužel právě tyto typy služeb jsou v reálných řešeních IS velice nevhodně implementovány, pokud vůbec. Špatný stav těchto služeb nejlépe ilustrují současné české knihovny. Proč knihovny kupují spousty bibliografických databází, když pak nejsou schopny uživatelům dodat příslušné primární dokumenty? Nákup sekundárních informačních zdrojů bez možnosti **efektivně a rychle získat popisované primární informace** má nulový efekt a představuje zcela nevhodně vynaložené finanční prostředky. Knihovny nabízejí informační pomůcky, která ale neslouží ke konečnému uspokojení uživatelských potřeb.

V rámci integrace informačních zdrojů bychom vždy měli dát uživateli možnost získat primární informaci, přestože k vyhledání uživatel použil pomocné informační zdroje, jako jsou např.

bibliografické databáze či katalogy. Primární informaci mu musí IS nabídnout přímo nebo alespoň zprostředkovat její získání, bez nutnosti použít jiný IS. U manuálně prováděných služeb je třeba zavádět kontrolní mechanismy, které kvalitu prováděných služeb průběžně kontrolují.

Nejběžnějšími typy služeb pro dodání primárních informací jsou následující:

- **přímé dodání dokumentů online** - uživatel získá dokument přímo z obrazovky v reálném čase v samoobslužném režimu bez účasti pracovníků knihovny, které tak odpadnou náklady za práci a zhotovení kopie
- **elektronické dodávání dokumentů (EDD)** – pracovník zprostředkuje manipulaci s klasickým fondem a provede digitalizaci požadované části dokumentu, čímž v procesu získávání vznikají dodatečné náklady za práci a zhotovení kopie. Tyto náklady jsou většinou přeneseny na vrub uživatele, navíc vzniká časová prodleva dodání dokumentu. Výhodou EDD je, že umí dodat dokumenty, které již nejsou v prodeji.
- **meziknihovní výpůjční služba** – zapůjčení fyzického exempláře dokumentu mezi knihovnami, který může uživatel využít většinou pouze prezenčně. Elektronická forma MVS má podobné rysy jako EDD, výraznou komplikací pro uživatele je však na rozdíl od EDD fakt, že MVS je vždy prováděna prostřednictvím konkrétní knihovny, tj. MVS si nemůže zajistit sám čtenář, těžko se proto bezešvě integruje do IS. Uživatel za tuto službu většinou platí a doba prodlevy dodání dokumentu je ještě delší než u EDD, často již na hranici únosnosti.
- **referenční služby typu „ptejte se knihovny“** – pracovník provozující organizace zodpoví volně položený dotaz za využití referenční příručky, klasického i elektronického fondu či internetu. Nejedná se tedy o běžné dodávání dokumentů, ale o faktografické rešerše, kdy uživatel získává rovnou požadovanou informaci. Tyto služby jsou často poskytovány velice rychle (mnohdy do 24 hodin) a v mnoha případech dokonce bezplatně.
- **odkaz na knihkupectví** – u dokumentů, které jsou stále ještě na knižním trhu, může IS přesměrovat uživatele na knihkupectví, kde si může uživatel požadovaný titul zakoupit. Díky elektronickým obchodům s podporou elektronických plateb jsou tyto nákupy čím dál oblíbenější, protože je mohou využívat uživatelé vzdáleně, a proto by měly být významnou alternativou při dodávání primárních dokumentů v IS. Stejně jako knihkupectví mohou při dodávání dokumentů posloužit i antikvariáty, čímž se nabídka rozšíří i o tituly, které již z knižního trhu zmizely.
- **odkaz na autora** – zajímavou možností, jak získat primární dokumenty či informace zejm. ve vědecké a technické oblasti, je kontakt na autory příslušných děl. Uživatel může např. elektronickou poštou požádat autora o zaslání kopie jeho článku či dokonce s ním popisovanou problematiku diskutovat.
- **odkaz na jiného vlastníka dokumentu** – často poslední možnost, jak primární dokument získat, je zkontaktovat jiného vlastníka příslušného dokumentu, který s integračním systémem nespolupracuje. Může to být např. knihovna, která má publikaci ve svém fondu, ale neprovozuje EDD ani žádné jiné automatizované služby, nebo účastníci odborných elektronických konferencí či maillistů. Jedinou možností je obrátit se na takového vlastníka a požádat ho o zapůjčení či jiné sdílení dokumentu, samozřejmě při dodržení autorských práv.

IS by měl na prvním místě nabízet takový **způsob** získání primárních informací, který je v konkrétním případě pro konkrétního uživatele **nejrychlejší a nejefektivnější**. Zároveň by měl mít uživatel vždy možnost volby, pokud existují při získávání alternativy. Např. u bibliografického záznamu, pro který nelze popisovaný dokument přímo plnotextově zobrazit na obrazovce, by měl mít uživatel volbu, jestli si dokument objedná přes EDD nebo v knihkupectví.

Knihovny a další organizace, které vlastní cenné informační zdroje v neelektronické podobě, by se měly snažit o jejich **digitalizaci**. Převedení informačních zdrojů do elektronické podoby a jejich zpřístupnění online v rámci IS představuje nejlepší řešení a eliminuje potřebu řešit dodávání primárních informací pomocí navazujících služeb.

### 2.8.3.1 Problémy spjaté s elektronickým dodáváním dokumentů v ČR

Elektronické dodávání dokumentů se v mnoha českých projektech potýká s velkou řadou problémů, které brání většímu využívání této služby.

Prvním z těchto problémů je **nedostatečný informační aparát**, který by uživateli umožňoval přesně určit, jakou část dokumentu si přeje doručit. Uživatel málokdy potřebuje elektronicky dodat celou monografii či sborník. EDD je tedy třeba integrovat s databázemi článků jako analytických jednotek a s citačními databázemi, které uvádějí i přesné rozsahy stran, aby uživatel nemusel kvůli objednavce navštívit knihovnu a určit rozsah z fyzického exempláře.

Dalším velkým problémem jsou **požadavky na ochranu autorských práv** z hlediska pořizování elektronických kopií dokumentů (zejm. dočasnost těchto kopií a zabránění dalšímu nekontrolovatelnému kopírování). Na některé typy dokumentů nelze kvůli autorským právům dokonce EDD aplikovat vůbec, např. notové záznamy. Legislativa se v této oblasti často mění, proto by měl IS umět reagovat na jakékoli změny. Existují technologie, které umí těmto požadavkům na ochranu vyhovět, např. simulací klasické výpůjčky dokumentu pomocí omezení časové platnosti a následné expirace elektronické kopie (více informací o technologiích k zajištění autorských práv – viz kapitola 3.3).

Problémem u služeb EDD v ČR je i to, že neexistuje **jednotný, centrálně řízený systém EDD**. EDD dnes provozuje prakticky každá větší knihovna, mnohdy na základě vlastního, proprietárního řešení. Uživatelé, kteří mají o EDD konkrétní knihovny zájem, se pro tento účel musí v knihovně zaregistrovat. Pokud integrační systém pokrývá fondy více knihoven provozujících EDD, musí se uživatel registrovat v každé knihovně zvlášť. Špatná kooperace mezi více systémy EDD je způsobena i tím, že není k dispozici jednoduchý a levný způsob, jak přijímat za služby **mikroplatby**.

Příkladem zahraničních řešení, kde je možné se inspirovat i použít z chyb, je britský EDD systém BLDSC (British Library Document Supply Center), německý systém SUBITO postavený na časopisecké databázi ZBD či systém JASON německých univerzitních knihoven.

#### 2.8.4 Problematika plateb

Informační zdroje a služby jsou nabízeny buď bezplatně nebo za úhradu. Důvodů ke zpoplatnění může být více: může je vyžadovat komerční provozovatel informačního systému, můžeme chtít pokrýt náklady na provozování nějaké služby (např. EDD), potřebujeme pokrýt úhradu kolektivnímu správci autorských práv apod. Potřebujeme-li od uživatelů prostřednictvím IS vybrat poplatek, často velice nízký, nastává vážný problém – problém **elektronických mikroplateb**.

Mikroplatba je EU definována jako platba do výše €15 a týká se většinou drobných plateb za služby na internetu. Malé částky, kterých se tyto transakce týkají, dávají malý prostor pro tvorbu zisku, proto řada mikroplatebních systémů balancuje na hranici pokrytí provozních nákladů. Předpokladem úspěšného mikroplatebního systému je velký obrat, který může být vzhledem k nízkým částkám dosažen pouze velkým množstvím provedených plateb.

Mikroplatebním systémům se v ČR příliš nedaří. Příčin můžeme vidět hned několik: uživatelé internetu si teprve pomalu zvykají za služby internetu platit, platební systémy mají vzhledem k nárokům na bezpečnost a spolehlivost vysoké provozní náklady a v neposlední řadě je to zákon o platebním styku 124/2002 Sb.<sup>56</sup>, který vešel v platnost 1. ledna 2003. Tento zákon stanovil, že výhradní právo provozovat podobné systémy mají pouze subjekty s bankovní licencí. Proto mnoho mikroplatebních systémů založených na virtuálních účtech a peněženkách muselo ukončit svoji činnost.<sup>1</sup>

Jako velmi nadějně se dnes jeví **mikroplatební systémy SMS**, protože téměř každý potencionální uživatel dnes vlastní mobilní telefon. Mobilní telefon jako platební nástroj je pro uživatele přijatelný a budí tolik potřebnou důvěru. Tento fakt potvrzují mnohé masové akce, kde se mobil jako mikroplatební nástroj využíval (např. hudební soutěž Superstar, volba Miss ČR, charitativní sbírky atd.). Mobilní operátoři si tuto příležitost uvědomují a prostřednictvím společné Asociace pro mobilní platby (Mobile Payment Association) usilují o zvýšení povědomí o mobilních platbách, vyvíjejí potřebné technologie a připravují vhodné podnikatelské modely. V současné době má i tento způsob plateb svoje nevýhody, zejm. poměrně vysoké poplatky, které si strhávají operátoři mobilních služeb, a značně nízký limit jednorázové platby.

Některé IS, provozující služby za úhradu, se vydaly cestou **uživatelských účtů se složeným kreditem**. Uživatel si formou složeného kreditu předplatí portfolio služeb a postupně z něho



odčerpává poplatky za úhradu konkrétních služeb. Uživatelé mohou kredit navyšovat příkazem k úhradě, poštovní poukázkou či složením hotovosti na pokladně provozující organizace. Kredit může být uložen na účtu konkrétního IS nebo v elektronické peněžence (většinou v podobě chipové karty). Pokud se dohodne více subjektů na spolupráci, je možné využívat jeden kreditní mikroplatební systém pro služby více IS a subjektů. Příklady takové spolupráce vidíme už dnes na vysokých školách (studentské karty) či v rámci měst (tzv. městské karty). Výhodou kreditního způsobu je nezávislost na službách bank či zprostředkovatelů (např. mobilních operátorů).

## 2.9 VÝZNAM STATISTIK PRO INTEGRACI A VYUŽITÍ INFORMAČNÍCH ZDROJŮ

U každého informačního systému je třeba pečlivě sledovat všechny jeho procesy, analyzovat je a vyvozovat z nich opatření pro další provoz IS. U integračního systému je tato potřeba ještě silnější, protože procesy v něm jsou různorodé a vznikají mezi nimi složité vztahy. Souhrn všech monitorovacích funkcí, ať již na technické, uživatelské nebo jiné úrovni, a z nich odvozovaných analýz nazýváme **zpětnou vazbou**.

Nejzákladnějším prostředkem zpětné vazby jsou bezesporu statistiky a v této kapitole jim budeme věnovat pozornost z hlediska technického provozu IS, z hlediska chování uživatelů, z hlediska akvizice informačních zdrojů, z ekonomického hlediska a z pohledu dalšího vývoje projektu IS. Kromě toho mohou statistiky IS sloužit jako cenná data pro širší statistické analýzy organizace, jako je např. měření výkonu a kvality v knihovnách.

Národní knihovna ČR, která sbírá a zpracovává statistická data českých knihoven, používá kombinaci mezinárodních normy **ISO 2789** Dokumentace a informace – Mezinárodní knihovnická statistika, která specifikuje pravidla pro sběr a zpracování statistik knihovních a informačních služeb, a standardu iniciativy **COUNTER** (Counting Online Usage of Networked Electronic Resources) pro měření využívání síťových elektronických zdrojů. Mezinárodní platnost normy ISO 2789 umožňuje srovnání úrovně informačních služeb mezi mnoha organizacemi a určuje, jaké data je třeba sbírat pro potřeby normy **ISO 11620** Dokumentace a informace – Ukazatele výkonnosti knihoven. Sledování specifikovaných ukazatelů je navíc předpokladem splnění požadavků normy pro řízení jakosti ISO 9001.

ISO 2789 definuje 4 stěžejní ukazatele, které je třeba sledovat u každého informačního zdroje/slужby:

- počet připojení
- počet přejímaných dokumentů
- počet přejímaných záznamů
- počet virtuálních návštěv

Vedle toho byly definovány následující doplňující ukazatele, jako nepovinné ale vhodné:

- doba připojení
- počet odmítnutých připojení
- počet vyhledávání (dotazů)
- počet připojení na internet

### 2.9.1 Sledování IS z technických důvodů

Cílem sledování statistik IS z technických důvodů je udržení optimálního chodu IS, případně pomoc a kontrola při nasazování nové části či komponenty IS (tzv. debugging). Pečlivou analýzou sledovaných ukazatelů můžeme odhalit příčinu chyb v IS či dokonce zaznamenat včas varovné signály (např. blížící se zaplnění diskové kapacity) a projevu chyb předejít. Analýzou chování integrovaných informačních zdrojů také odhalujeme problémy a výpadky na vzdálených serverech, které pak můžeme řešit s jejich správci.

Jako vstupní data statistik pro technický monitoring slouží zejm. generované **logy, reporty a alarmy** z IS nebo jeho vrstev (např. logy databázového stroje, logy aplikačního serveru apod.).

### 2.9.2 Sledování chování uživatelů

Analýzou statistických ukazatelů můžeme popsat, jakým způsobem uživatelé s IS pracují a odhalovat tak mnoho nedostatků uživatelského rozhraní. Vidíme, jaké funkce uživatelé volají nejčastěji, jaké naopak nevyužívají, v jaké fázi pracovního postupu se zarazí a dále nepokračují

(důsledek neintuitivních komponent rozhraní) a z toho odvozovat příčiny a provádět opatření, zejm. změnou uživatelského rozhraní, změnou nabídky funkcí či zdrojů. Důležitým zdrojem informací je i analýza dotazů přicházejících na helpdesk IS či sběrem požadavků typu „co chcete zlepšit“. Sledování chování uživatelů uvnitř IS můžeme doplňovat i o další způsoby zjišťování problémů a potřeb uživatelů, např. ankety nebo dotazníkové průzkumy, nicméně ty již přinášejí riziko subjektivity a nepřesnosti.

Zdrojem dat pro analýzu uživatelského chování a uživatelských potřeb jsou zejm. **logy generované uživatelským rozhraním** (např. logy webového serveru), seznamy stížností, dotazů a požadavků na **helpdesk** a různé metody klasického statistického průzkumu.

### 2.9.3 Sledování využití zdrojů z akvizičních důvodů

Statistická analýza využití jednotlivých informačních zdrojů (zejm. kvantitativní charakteristika) přináší jedny z nejcennějších informací pro proces akvizice. Nabídku informačních zdrojů v integrovaném systému je třeba sledovat z hlediska pokrytí i obsahu (rozsah, stáří, zaměření) a z hlediska využití (efektivita z hlediska výtěžnosti fondu). Tyto ukazatele můžeme využít i jako vstupní data (administrativní metadata) pro ERM systémy, které na základě nich dokáží popisovat chování každého informačního zdroje (tuto problematiku řeší již dříve zmíněné sdružení Digital Library Federation v rámci své iniciativy Electronic Resources Management Initiative).

U zdrojů, které jsou intenzivně využívány, můžeme uvažovat o rozšíření podobně zaměřených zdrojů či je obohatit o navazující služby, které zvýší uživatelský komfort při jejich využívání. Zdroje, které se téměř vůbec nevyužívají, indikují, že mohou být staženy z nabídky a v případě komerčních informačních zdrojů je to jasný signál k tomu, aby nebyla obnovována licence. Důvodem malého využití však může být i nevhodná implementace, integrace či propagace, pak je třeba tento problém včas rozpoznat a provést nápravná opatření.

Integrace nových zdrojů by se měla řídit analýzou požadavků uživatelů (průzkum) a analýzou dotazů, které vracely nulové nebo naopak příliš velké výsledky. Nulový výsledek vyhledávání znamená, že na dané téma nemáme k dispozici potřebné informační zdroje, naopak příliš velké množství výsledků znamená, že dané téma je pokryto pouze univerzálně a máme nedostatek úzce zaměřených informačních zdrojů.

Vstupními daty pro tyto analýzy je zejm. **historie vyhledávacích dotazů** s údaji o cílových informačních zdrojích.

### 2.9.4 Sledování využití zdrojů z ekonomických důvodů

S akvizicí souvisí i ekonomické hledisko sledování **výtěžnosti informačních zdrojů** a opět jsou nejpreciznějším zdrojem informací statistiky. Vždy je třeba pečlivě zvažovat, jaké zdroje se vyplatí koupit trvale, které jen pronajímat (např. nákupem časově omezených licencí) a které uživatelům zajišťovat pouze v rámci MVS nebo jako objednávku na vyžádání. Finanční rozdíly mezi těmito způsoby zajišťování informačních zdrojů jsou často ohromné.

Pokud si informační management organizace nechá zpracovat analýzu výtěžnosti nakupovaných zdrojů na základě reálných statistik, jistě dojde k řadě úsporných opatření, které při zachování výběru informačních zdrojů ušetří nemalé finanční prostředky. Statistické podklady zároveň slouží ke zdůvodnění a obhájení rozpočtu knihoven, zejm. na nákup informačních zdrojů, které uživatele hojně využívají, případně na tvorbu databází digitálních objektů. Měl by proto existovat technický nástroj, jak z celkových statistických údajů připravit report pro konkrétní informační zdroj nebo konkrétní spolupracující instituci nebo provozovatele.

### 2.9.5 Zpětná vazba pro další vývoj projektu

V předchozích kapitolách jsme několikrát zmiňovali, že každý IS prochází svým **životním cyklem**. Pokud chceme, aby IS pro integraci informačních zdrojů co nejdéle a co nejefektivněji pracoval a uspokojoval informační potřeby uživatelů, je třeba reagovat na měnící se požadavky a očekávání uživatelů. Typicky se toto děje formou zavádění **nových verzí IS**. Každá změna verze představuje samostatnou etapu životního cyklu a měla by reagovat na informace získané

zpětnou vazbou z verze předchozí. Statistiky získávané z provozu IS tvoří velmi významnou druh zpětné vazby pro jeho další vývoj.

Je však důležité sledovat chování nejen v rámci vlastního IS ale i jeho okolí. Užitečné je analyzovat vývoj podobných projektů, srovnávat konkurenční řešení, sledovat technologické a metodické trendy atd. Všechny tyto analýzy by se měly průběžně odrážet v souhrnném reportu, který bude jasně a přesně formulovat situaci, aby mohlo vedení projektu okamžitě reagovat na přicházející hrozby, eliminovat slabé stránky, využívat silné stránky a pohotově reagovat na nové trendy a příležitosti. Jedině tak může být projekt IS udržitelný a životaschopný.

## 2.10 PORTÁL JAKO ZASTŘEŠUJÍCÍ NÁSTROJ INTEGRACE

Při snaze o co nejtěsnější integraci informačních zdrojů a navazujících služeb nutně dojdeme k potřebě zastřešit všechny funkce a dílčí komponenty IS jedním jediným rozhraním, které by vystupovalo jako centrální uzel přístupu ke všem informacím a funkcím. Takovou technologií je bezesporu **portálové řešení informačního systému**. Výraz „portál“ je bohužel termín, který je v různých kontextech chápán zcela odlišně, proto je nutné nejprve specifikovat, jaký význam tohoto termínu zde máme na mysli. Výraz „information portal“ (informační brána) je natolik obecný, že je otázka, zda ho vůbec můžeme považovat za odborný termín. Bohužel se tento výraz již natolik prosadil a rozšířil do praktických technologických řešení a projektů, že ho lze těžko nahradit. Informační portál pro integraci informačních zdrojů se v mnohém odlišuje od vlastností a pojetí podnikových portálů, kterým se věnuje většina portálové literatury současnosti. Hlavní rozdíl lze spatřit v tom, že komunikace v podnikově orientovaných portálech je výrazně obousměrná (vzájemná komunikace mezi zákazníky, zaměstnanci, obchodními partnery a podnikem), kdežto v portálech orientovaných na využití informačních zdrojů a informačních služeb převažuje **tok informací směrem k uživateli**. Samozřejmě však není interakce s uživateli vyloučena, je naopak velice žádoucí. V každém případě ho musíme chápat v obecné rovině síťového informačního prostředí, které se vyznačuje velkou variabilitou a různorodostí.

### 2.10.1 Co je informační portál ve smyslu integrace informačních zdrojů

Pracovní skupina LCPAIG (Library of Congress Portals Applications Issues Group) definuje portál jako nástroj pro pokročilé získávání znalostí, který by měl být schopen:

- pomáhat uživateli vybrat relevantní informační zdroje
- pomáhat uživateli rozpoznat, které zdroje jsou pro jeho řešerši nejhodnější pomocí tvorby vyhledávacích skupin a bohatého popisu informačních zdrojů
- umožňovat paralelní vyhledávání popisných metadat z více heterogenních zdrojů, kam patří např. knihovní katalogy, komerční a licencované elektronické zdroje, databáze, webové stránky atd.
- pomáhat uživateli sestavit řešeršní dotaz a řídit proces vyhledávání
- zajistit, že vyhledané výsledky budou řádně deduplikovány
- převádět výsledky do čitelného formátu a umožnit uživateli výsledky interpretovat a manipulovat s nimi
- ukládat a exportovat výsledky pomocí různých výstupů (tisk, e-mail, stažení záznamů)
- linkovat výsledky na plné texty či jiné digitální objekty nebo na služby dodávání obsahu
- řízeného přístupu k informačním zdrojům a funkcím portálu pro autentikované uživatele na základě různých skupin a rolí.<sup>41</sup>

Tato definice je příliš úzce zaměřena na knihovní portály, proto ji poněkud rozšíříme, aby byla vhodná pro jakýkoli portál integrující informační zdroje:

Portál je nástroj, který poskytuje personalizovaný, bezpečný a unifikovaný přístup k jednotlivým částem portálu (tzv. portletům) a registrovaným uživatelům zajišťuje informační podporu ve všech aspektech jejich studijních, pedagogických, vědeckých a dalších činností. Portlety mohou být interní či externí a mohou zpřístupňovat lokální nebo vzdálené informační zdroje (knihy, časopisy, databáze, webové stránky, digitální sbírky, výukové materiály, IS školy atd.), informační služby (rezervace dokumentů, registrace, objednávky EDD, tvorba avíz, RSS atd.) a nástroje pro spolupráci (e-mail, kalendář, chat, blogy atd.).

Z uvedené definice vyplývají hlavní funkční oblasti portálového řešení:

- integrovaný přístup k informacím a službám a unifikovaný pohled na všechna data
- nabídka portletů podle potřeb uživatele a samoobslužná úprava jejich uspořádání/nastavení
- bezpečný a řízený přístup, vysoká dostupnost, škálovatelnost a otevřenost

Velký důraz je v portálech kladen na **samoobslužnost služeb** a na možnost nastavit si pracovní prostředí **podle svých potřeb**. Každý uživatel má jiné potřeby - má určité profesní, studijní a soukromé zájmy, určitou jazykovou vybavenost, pracovní návyky apod. Ideálně by měl portálový systém umožňovat uživateli přetvořit výchozí nastavení tak, aby měl v nabídce přesně ty služby, o které má zájem, a do určité míry mohl ovlivnit, jakým způsobem se budou chovat. Uživatel se v takovém případě spolupodílí na dotvoření svého pracovního prostředí a vzniká tak pro každého registrovaného uživatele **vlastní portál**. Typicky může uživatel vybírat informační zdroje, které se mu budou nabízet, nastavovat formy vyhledávání, způsoby zobrazování, může měnit nabídku služeb apod. Portálové řešení jako celek vytváří výchozí systémové prostředí pro uživatele.

Jinými slovy, portálový systém tvoří systémovou kostru (framework), na které je pomocí portletů budováno konkrétní portálové řešení. Hlavním přínosem portálového pojetí IS je unifikace vzhledu a chování všech částí IS, což uživateli zajišťuje zcela jednotné pracovní prostředí, které však může lehce kustomizovat.

### 2.10.2 Co je portlet

Základní funkční a aplikační jednotkou portálového řešení jsou portlety. Můžeme na ně pohlížet jako na samostatné funkční moduly, které jsou však schopny bezešvě spolupracovat se systémovou kostrou portálu a často i mezi sebou. Nezávislost portletů se projevuje také tím, že lze do portálu libovolně přidávat nebo z portálu odebírat, i když existují i skupiny portletů, které na sobě závisejí. Z funkčního hlediska zajišťují portlety standardizovaným a bezpečným způsobem specifické služby a zpřístupňují informační zdroje. Portálový systém zajišťuje dílčím portletům potřebné systémové prostředí, např. autentikaci a vazby na adresářové služby, řízení přístupů k objektům, unifikovaný vzhled a zobrazení, systémové vstupy a výstupy, datové konektory, podporu pracovních postupů apod.

Pokud to portálové řešení umožňuje, mohou si sami uživatelé vybírat z nabídky portletů (z tzv. repozitáře portletů) a volitelně si je umisťovat do svého uživatelského portálu. Funkcionalita portálu je v takovém případě zcela dynamická a přizpůsobitelná měnícím se potřebám uživatele.

V portálech pro integraci informačních zdrojů můžeme nabízet typicky následující portlety:

- paralelní vyhledávač (klíčová komponenta)
- objednávkový systém pro EDD
- čtenářská schránka AKS domovské knihovny
- personalizovaná úložiště pro dokumenty, záznamy, dotazy, SDI atd.
- statické informační stránky portálu, fotogalerie, recenze atd., snadno editovatelné přes vizuální publikační systém
- obsah jiných IS přebíraný přes RSS
- kontakty nebo adresáře
- prostor pro publikování samotnými uživateli (např. blogy, wiki)
- nápovědy, průvodce, referenční příručky
- interakce s organizací (návrhy na akvizici, chat s knihovníkem, ankety, elektronické formuláře atd.)
- možnost komentovat vyhledané informační objekty formou poznámek či taggování (přirazování řízených či neřízených klíčových slov) a sdílet je s ostatními uživateli

Nabídka se může v praktickém projektu lišit, záleží na informačním prostředí organizace, na potřebách uživatelů a v neposlední řadě i na snaze tvůrců portálu přinášet zajímavé a užitečné služby.

### **2.10.3 Informační portál a možnost spolupodílet se na vytváření obsahu**

Webové systémy v posledních letech směřují jednoznačně k naplnění touhy uživatelů podílet se na vytváření informací v IS dostupných na webu. Tyto tendence jsou obecně označovány jako Web 2.0. Informační portál umožňuje díky své komponentové struktuře tyto potřeby naplnit.

Do portálu můžeme formou modulů a portletů přidávat funkce, které registrovaným uživatelům IS dovolí sdílet své názory formou blogů nebo komentářů, diskutovat mezi sebou na daná témata, využívat služeb IS na svých mobilních zařízeních, libovolně si sestavovat a kombinovat RSS kanály podle momentálních potřeb, vytvářet v otevřených komunitách informační obsah formou wiki, hodnotit obsah a kvalitu informačních zdrojů i jednotlivých objektů formou ratingu nebo uživatelských klíčových slov (tzv. folksonomie) atd. Možnosti jsou široké. Organizace provozující informační portál by měla neustále mapovat a analyzovat potřeby uživatelů a všeobecné trendy, aby trvale udržitelným rozvojem zachovala nebo dokonce zvyšovala návštěvnost portálu.

## **2.11 SHRNU TÍ**

**Architektura IS pro integraci informačních zdrojů je vždy odrazem konkrétních projektových řešení. Můžeme však najít mnoho prvků, které jsou pro tato řešení společná. Návrh architektury IS a její postupné ladění by mělo vždy vycházet z pečlivě prováděných projektových fází podle zvolené metodiky. Integrace informačních zdrojů se vyznačuje výjimečnými specifiky, zejm. vysokou nezávislostí externích zdrojů, které provozují jiné, na nás zcela nezávislé subjekty. Z toho důvodu je často nemožné použít čistě centrálně pojatá integrační řešení, ale systém jsme nuceni distribuovat mezi centrální prvek a vzdálené komponenty. Důraz je proto kladen na unifikaci heterogenních prvků, zejm. na sjednocení uživatelských rozhraní a dotazovacích jazyků, spolehlivou autentikaci, bezpečnost a statistiky využití s ohledem na požadavky jednotlivých provozovatelů informačních zdrojů. Distribuované prostředí přináší i řadu specifických problémů k řešení, zejm. v oblasti vyhledávání. Klíčovým momentem je schopnost architektury IS napojit navazující služby na informace vrácené z informačních zdrojů. IS pro integraci informačních zdrojů mohou být úspěšně začleněny do portálových systémů, čímž uživatelé získají nejkomplexnější nabídku pracovních nástrojů, mezi nimiž hrají integrované informační zdroje knihoven významnou roli.**

### 3 TECHNICKÉ NÁSTROJE INTEGRACE INFORMAČNÍCH ZDROJŮ

V této kapitole si ukážeme některé technologie, které lze využít při integraci informačních zdrojů. Není možné zde zmínit, natož podrobně popsat všechny dostupné technologie. Záměrně jsou zde vynechány konkrétní produkty jednotlivých výrobců software. Důvodem je velká proměnlivost nabídky a snaha neprosazovat konkrétní řešení. Chybí zde proto přehled vyhledávacích a databázových strojů, metavyhledávačů, portálových řešení a dalších informačních systémů, přestože mohou v konkrétních projektech vystupovat jako dílčí komponenty zajišťující svěřené funkce. Dnešní dynamické prostředí mnohdy vyžaduje modulová řešení (tzv. suity). Jsou to skládky sdílených a integrovaných komponent různých výrobců, protože ve většině případů není možné najít produkt, který by beze zbytku splňoval všechny požadavky projektu.

#### 3.1 VÝZNAM TECHNICKÝCH STANDARDŮ

Jedním ze základních předpokladů integrace heterogenních informačních zdrojů a služeb je schopnost vzájemné spolupráce neboli interoperabilita. Schopnost spolupráce v technickém smyslu nevzniká náhodně, ale je definována tím, nakolik dané technologie či služby odpovídají dohodnutým specifikacím. Souborům specifikací pro interoperabilitu říkáme standardy. Standardy musí být dostatečně volné, aby vyhovovaly všem zúčastněným, ale dost přísné, aby zajistily potřebnou kvalitu. Standardy musí být schváleny a přijaty buď velkou skupinou zainteresovaných stran (např. výrobců hardware) nebo národní či mezinárodní standardizační organizací, jako jsou International Organization for Standardization (ISO), American National Standards Institute (ANSI), European Committee for Standardization (CEN), InterNational Committee for Information Technology Standards (INCITS) či World Wide Web Consortium (W3C).

Některé standardy mohou být patentovány a chráněny, naopak některé jsou zcela volné a mohou být využity kýmkoli, kdo se pro ně rozhodne. Takovými volnými standardy říkáme **otevřené standardy** a právě ony jsou nejvýznamnějšími hybateli technologického pokroku v systémové integraci.

Opakem otevřených standardů jsou tzv. **proprietární technologie**, které vznikají většinou v rámci jediného výrobce nebo jednoho projektu šité na míru. Proprietární technologie je uzavřená, vyznačuje se výraznými privátními specifiky, které brání bezševé komunikaci a propojování s okolím.

Při projektování IS pro integraci informačních zdrojů stojíme vždy před základní otázkou, zda vyvinout proprietární řešení nebo využít technologie postavené na otevřených standardech. Proprietární technologie může být pro konkrétní situaci vhodnější, může být i levnější a krátkodobě efektivnější. Existují však vážné důvody, proč se přiklánět ke standardizovaným řešením:

- **nezávislost na technologii, produktu či službě jediného výrobce** – výrobce může zaniknout, ukončit podporu dané technologie či znevýhodnit obchodní podmínky
- **otevřené standardy umožňují svobodně vybírat jednotlivé komponenty a kombinovat je** – v integračním řešení můžeme používat dílčí technologie různých výrobců a propojovat je mezi sebou, čímž můžeme využít těch nejlepších vlastností
- **standardy v sobě obvykle odrážejí nejlepší možná řešení** – standardy vznikají na základě zkušeností odborníků v dané oblasti, před schválením jsou dlouhou dobu připomínkovány – pokud neexistují vážné důvody k proprietárnímu řešení, je nejjednodušší používat standardy, které jsou zárukou kompatibility s dalšími prvky v okolí, navíc nemusíme vynakládat finance na vývoj vlastních řešení
- **standardy jsou postupně vyvíjeny** – vznikající nové verze otevřených standardů jsou impulsem k progresivním změnám řešení u všech výrobců bez nutnosti přímé koordinace

Pokud již IS nepoužívá standardizované technologie na všech vrstvách, měl by být alespoň vybaven standardizovanými vstupy a výstupy, aby mohl systém spolupracovat s jinými IS.

Příkladem může být knihovnický systém, který sice ukládá záznamy ve svém vlastním vnitřním formátu, ale na úrovni funkcí pro import a export dokáže zpracovat záznam v MARC 21.

Následující tabulka shrnuje nejvýznamnějšími otevřené standardy používané v IT v knihovnách:

Komunikační protokoly	HTTP, Z39.50, FTP
Protokoly speciálních služeb	OAI-PMH, openURL, EDI, SRU/SRW
Dotazovací jazyky	CCL, CQL, Z39.50, SQL
Formáty popisných dat	MARC 21/MARCXML, DC, MODS, MADS, METS
Předmětové systémy	Konspekt, MDT, DDT
Formáty plných textů	HTML/CSS, PDF, RTF
Formáty grafických objektů	JPEG, PNG, TIFF, DjVu, SVG
Formáty zvukových objektů	WAV, MP3, AU, MIDI
Formáty video objektů	AVI, MPG, DVD
Autentikace	LDAP, EAP, Shibboleth, SAML

Základní popis uvedených standardů je uveden ve slovníku zkratk (viz kapitola 8).

## 3.2 TECHNOLOGIE PRO VYHLEDÁVÁNÍ A KOMUNIKACI SE VZDÁLENÝMI INFORMAČNÍMI ZDROJI

V kapitole 2.7.1.1 jsme mluvili o centrálně a distribuovaně pojatých způsobech vyhledávání. Oba dva způsoby mají své výhody i nevýhody a jejich nasazení je často podmíněno nejen technickými požadavky, ale i politikami spolupracujících subjektů. Díky internetizaci knihoven, která zajistila většině knihoven permanentní připojení na internet, se v minulých letech výrazně prosadily technologie distribuovaného vyhledávání, zejm. protokol Z39.50. V současné době je naopak cítit jistá tendence návratu k centralizovaným typům vyhledávání, což je způsobeno rozvojem elektronických archivů a digitálních repozitářů, které používají ke spolupráci s okolím protokol OAI-PMH. Protokol Z39.50 se navíc může zdát z dnešního pohledu poněkud zastaralý, protože na rozdíl od modernějších technologií neinklinuje k webovému pojetí služeb, které jsou postaveny na komunikačním protokolu HTTP a formátu XML. Standard SRU/SRW, který byl vyvinut jako XML nástupce protokolu Z39.50, se zatím v praxi příliš nerozšířil, nicméně mnoho vývojářů v něm vidí cestu dalšího vývoje.

Podívejme se proto na tyto protokoly podrobněji.

### 3.2.1 Z39.50

Komunikační a aplikační protokol Z39.50 pro vyhledávání a přejímání dat jsme v této práci již mnohokrát zmiňovali. V současné době totiž představuje nejrozšířenější technologii pro distribuované vyhledávání v knihovnických aplikacích a informačních zdrojích. Hlavní výhoda Z39.50 spočívá v tom, že nabízí zcela **nezávislou abstraktní vrstvu**, která umožňuje propojit v podstatě jakékoli informační zdroje mezi sebou či s integračním prvkem, a to nezávisle na jejich operačních systémech a na jejich databázové a aplikační vrstvě. Vzhledem ke svému masivnímu rozšíření a širokým možnostem využití nemá protokol Z39.50 pro distribuované vyhledávání v současné době konkurenci.

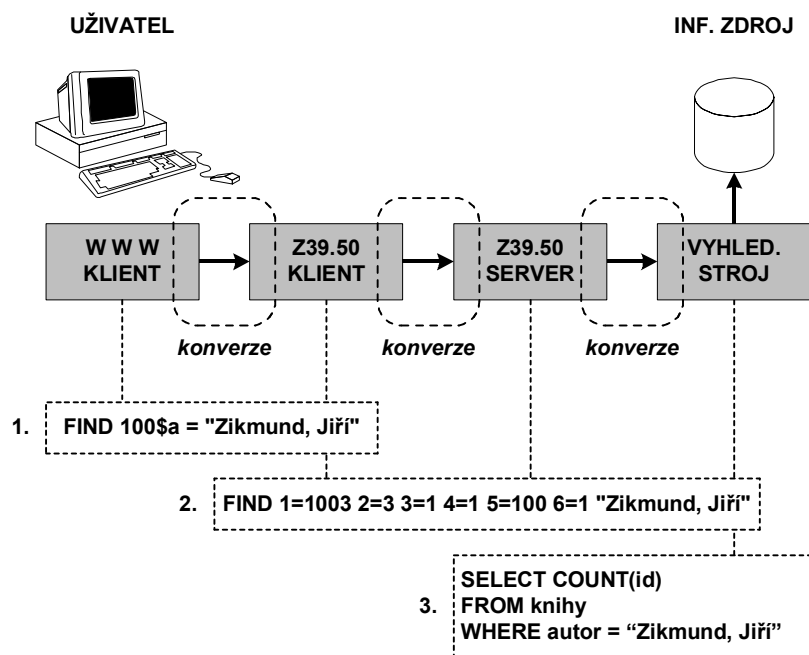
Aktuální verze Z39.50 pochází z roku 2003 (označována jako Z39.50-2003). V praxi je však nejrozšířenější starší verze **Z39.50-1995**, která byla v roce 1997 normalizována jako ISO 23950. Standard Z39.50 definuje kromě komunikační a aplikační funkcionality také vlastní dotazovací jazyk, takže zajišťuje nezávisle na jiných technologiích všechny funkce pro vyhledávání. Jádrem standardu Z39.50-1995 je definice 6 typů dotazů a 6 různých množin datových atributů. Pro oblast knihoven je nejvyužívanější typ dotazů Type-1 (kde jsou termy kombinované booleovskými operátory) a množina atributů Bib-1 (odpovídající bibliografické struktuře záznamů). Protože specifikace protokolu Z39.50 je poměrně rozsáhlá a otevřená, je třeba ke vzájemné spolupráci definovat tzv. **profily Z39.50**, které říkají, jaké atributy a jaké možné hodnoty budou v konkrétních projektech minimálně podporovány. Součástí profilu mohou

být také předepsané způsoby indexování jednotlivých polí, která jsou na Z39.50 atributy mapovány, aby všechny zúčastněné IS vracely srovnatelné výsledky.

Při komunikaci Z39.50 vystupuje informační zdroj jako Z39.50 server a prvek, který ho využívá, vystupuje jako Z39.50 klient. Komunikace mezi klientem a serverem je přesně specifikována protokolem, který kromě vyhledávání podporuje i další operace, včetně řazení, editace či mazání dat. Komunikace vypadá zjednodušeně takto:

- Systém, který vystupuje jako Z39.50 klient, nejprve převede konkrétní dotaz do syntaxe Z39.50 na základě mapování (např. z formátu MARC 21 nebo z parametrů zadaných uživatelem v uživatelském rozhraní). Dotaz je ve výsledku složen ze seznamu databází, ve kterých má být hledání provedeno, a z vyhledávacích výrazů (termů) s uvedenými hodnotami Z39.50 atributů pospojovaných booleovskými operátory.
- Klient se následně spojí se serverem a po úvodní výměně parametrů pošle na server dotaz Z39.50.
- Server dotaz přijme a následně ho přemapuje do interní syntaxe informačního zdroje, např. do SQL. Poté provede vyhledávání v příslušném systému správy bází dat.
- Vrácená množina výsledků je převedena do formátu, který klient očekává, a je odeslána buď jako celek nebo po částech zpět klientovi.
- Klient vrácené výsledky zpracuje a použije v daném kontextu na výstupu.

Na následujícím obrázku je příklad Z39.50 komunikace. Dotaz na pole 100\$a formátu MARC 21 s výrazem „Zikmund, Jiří“ je přemapován do syntaxe Z39.50 s typem dotazu Type-1 a množinou atributů Bib-1 specifikujícími použitá pole, strukturu, zarovnání atd. Server, který dotaz přijme, provede konverzi do syntaxe jazyka SQL a následně provede vyhledávání v databázi.



Obr. 17 – Komunikace Z39.50

Přesnou specifikaci normy Z39.50 lze získat na stránkách Knihovny amerického Kongresu<sup>54</sup>, která se prostřednictvím agentury „Z39.50 Maintenance Agency“ stará o další vývoj standardu Z39.50 a poskytuje potřebnou podporu. V ČR se o propagaci a koordinaci využívání Z39.50 stará sdružení ZIG-CZ (Z39.50 Implementors Group)<sup>53</sup>, které zaštiťuje Státní technická knihovna. Toto sdružení v únoru 2004 vyhlásilo ve spolupráci s Národní knihovnou ČR **Z39.50 profil JIB**<sup>55</sup>, který vymezuje minimální požadavky na nastavení Z39.50 atributů u českých katalogů a bibliografických databází pro účely jednoduchého vyhledávání.

Ve snaze přizpůsobit protokol Z39.50 modernímu webovému pojetí služeb, zjednodušit ho a zachovat všechny jeho osvědčené výhody, vznikla iniciativa Z39.50 International - Next



Generation (ZING), která připravila dva nové standardy SRU a SRW. Oproti Z39.50 představují SRU/SRW často levnější a univerzálnější řešení, zejm. s ohledem na vyhledávání v prostředí WWW, je však otázkou, nakolik je budou producenti AKS a databází ochotni implementovat do svých IS.

### 3.2.2 SRU/SRW

Informační technologie, na kterých je založen provoz informačních zdrojů, v poslední době směřují k maximálnímu zjednodušení a unifikaci. Prosazují se technologie, které se již dříve dobře osvědčily u služby WWW, tedy komunikační protokol HTTP a značkovací jazyky, především XML. Ukázalo se, že kombinace HTTP/XML je natolik univerzální a jednoduchá, že na ní lze postavit obecně jakékoli informační služby. Tento trend neminul ani protokol Z39.50. V listopadu 2002 vyvrcholila snaha již zmíněné iniciativy ZING a vznikla první verze protokolů SRU a SRW - nových standardů pro distribuované vyhledávání v informačních zdrojích. V současné době jsou k dispozici ve verzi 1.1.

**SRU** (Search/Retrieve via URL) je vyhledávací protokol, který zajišťuje vyhledávání a získávání dat z informačních zdrojů v prostředí WWW. Dotaz vyjadřuje standardizovaným dotazovacím jazykem CQL (Common Query Language) a kóduje ho do URL. Myšlenka SRU vychází ze starší technologie openURL, která taktéž k zasílání bibliografických údajů využívá URL. URL se skládá z adresy SRU serveru, který přijímá požadavky, a z části, která obsahuje samotný CQL dotaz v předepsané syntaxi. Adresa a vyhledávací část jsou odděleny znakem „?“, jednotlivé parametry dotazu potom znakem „&“. Parametry jsou tvořeny dvojicemi „klíč=hodnota“, kde klíče předepisuje norma SRU. Podporovány jsou tři základní operace: explain (pro získávání informací o databázi a podporovaných vlastnostech), searchRetrieve (zajišťující vlastní vyhledávání) a scan (k procházení rejstříku databáze). SRU umožňuje zasílat data v kódování Unicode, což zajišťuje jeho univerzálnost. Jednoduchý SRU dotaz může vypadat např. takto:

```
http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=olomouc
```

Výsledky jsou ze serveru vráceny ve formátu XML v požadovaném schématu (MARCXML, Dublin Core, MODS a další) a obsahují údaje v závislosti na volané operaci. Kromě vyhledávání podporuje SRU také řazení výsledků, specifikaci CSS stylu pro zobrazení, autentikaci a další podpůrné funkce. Celý SRU dotaz lze posílat zcela standardně metodou GET nebo POST jako běžné URL, takže nevznikají žádné komplikace při síťové komunikaci.

**SRW** (Search/Retrieve Web Service) je obdobou SRU. Komunikace však není realizována prostřednictvím URL, ale pomocí XML zpráv vyměňovaných mezi klientem a serverem prostřednictvím protokolu HTTP. Způsob prezentace XML zpráv je definován standardem webových služeb SOAP (Simple Object Access Protocol) podle agentury W3C. Z toho vyplývá, že jak SRW server, tak i SRW klient musí podporovat SOAP, konkrétně verzi minimálně 1.1. Syntax dotazu i odpovědi je stejný jako u SRU, s výjimkou názvů operací a možností volat CSS styl. Dotaz v podobě SRW může vypadat takto:

```
<SOAP:Envelope xmlns:SOAP="http://schemas.xmlsoap.org/soap/envelope/">
  <SOAP:Body>
    <SRW:searchRetrieveRequest xmlns:SRW="http://www.loc.gov/zing/srw/">
      <SRW:version>1.1</SRW:version>
      <SRW:query>(dc.author exact "svoboda" and dc.title >= "buzuluk")</SRW:query>
      <SRW:startRecord>1</SRW:startRecord>
      <SRW:maximumRecords>10</SRW:maximumRecords>
      <SRW:recordSchema>info:srw/schema/1/mods-v3.0</SRW:recordsSchema>
    </SRW:searchRetrieveRequest>
  </SOAP:Body>
</SOAP:Envelope>
```

Vývoj a podporu obou standardů zajišťuje agentura „SRU Maintenance Agency“ při Knihovně amerického Kongresu. Přestože je název jiný, jedná se ve skutečnosti o stejnou agenturu jako pro protokol Z39.50. Na stránkách agentury lze získat další informace i přesné specifikace standardů.

### 3.2.3 OAI-PMH

Zatímco protokoly Z39.50 a SRU/SRW slouží k prohledávání vzdálených informačních zdrojů, pomocí nichž můžeme budovat distribuované IS (např. virtuální souborné katalogy), protokol OAI-PMH reprezentuje zcela opačnou filosofii – je to nástroj, který umožňuje integračnímu systému získávat (sklízet) popisné údaje (metadata) z jednotlivých informačních zdrojů a centralizovaně je ukládat. V kontextu integrace informačních zdrojů umožňuje budování **fyzicky centralizovaných úložišť popisných údajů** informačních objektů a následné vytváření vyhledávacích indexů. Vyhledávání pak probíhá přímo v integračním IS na vybudovaných indexech, bez účasti vzdálených informačních zdrojů.

Protokol **OAI-PMH** (Open Archives Initiative Protocol for Metadata Harvesting) slouží pro sklizeň popisných dat, nikoli pro sklizeň popisovaných informačních objektů jako takových. Primární informační objekty, pokud v elektronické podobě existují, zůstávají v úložišti vzdáleného informačního zdroje a jsou přenášeny až na vyžádání na základě linkování z popisného záznamu.

Veškerá komunikace protokolem OAI-PMH probíhá formou XML zpráv, které jsou posílány prostřednictvím HTTP mezi klientem (tzv. harvesterem) a serverem (tzv. repozitářem). Každý metadatový objekt je v repozitáři opatřen jednoznačným identifikátorem, který musí odpovídat specifikaci URI. Tento identifikátor je celosvětově jedinečný, protože kromě ID samotného objektu je jeho součástí doménové jméno příslušného serveru (např. oai:nkp.cz:340076). Standard OAI-PMH dále specifikuje, že povinným metadatovým schématem je nekvalifikované Dublin Core. Kromě toho však samozřejmě dovoluje podporovat i další metadatová schémata.

Protokol OAI-PMH nabízí pro účely sklizeň šestici základních příkazů: **Identify** (slouží pro získání technických i popisných informací o repozitáři, které následně klient využije k nastavení dalších operací pro sklizeň), **ListMetadataFormats** (slouží k vypsání seznamu všech podporovaných metadatových schémat), **ListSets** (slouží k vypsání seznamu všech skupin záznamů, které mohou být v repozitáři vytvořeny), **GetRecord** (slouží k získání konkrétního záznamu v požadovaném formátu podle uvedeného jednoznačného identifikátoru), **ListIdentifiers** (vrací seznam všech identifikátorů záznamů, který lze omezit uvedeným schématem, skupiny či časového rozpětí) a **ListRecords** (zajišťuje vlastní sklizeň, tj. vrací celé metadatové záznamy).

První stabilní verze protokolu OAI-PMH je verze 2.0 zveřejněná v červnu 2002 po několikaletém vývoji této technologie v rámci iniciativy Open Archives. Z původního záměru vytvořit řešení pro kooperaci mezi elektronickými archivy vědeckých prací se postupně vyvinula technologie otevřeného přístupu k jakýmkoli elektronickým zdrojům v oblasti digitálních knihoven, elektronických archivů nebo bibliografických či plnotextových databází. Dnes prostřednictvím protokolu OAI-PMH nabízejí svá data stovky informačních zdrojů po celém světě. Vzniká velká řada projektů, které sklizená metadata využívají pro tvorbu citačních rejstříků nebo centrálních indexů virtuálních digitálních repozitářů, pro integrované prohledávání informačních databází společně s webovými zdroji či pro následné zpřístupnění přes Z39.50, kdy budované IS slouží jako dílčí agregátory.

## 3.3 TECHNOLOGIE NA ZAJIŠTĚNÍ OCHRANY AUTORSKÝCH PRÁV A LICENCÍ

Jak jsme již zmínili v kapitole 2.1.6, je u řady informačních zdrojů vyžadována ochrana autorských práv, dodržování licenčních podmínek a uplatňování speciálních režimů distribuce dokumentů. Provozovatelé, producenti, nakladatelé či autoři se v licenčních smlouvách zaměřují zejm. na kontrolu **pořizování kopií dokumentů**, kontrolu **modifikace dokumentů** a na kontrolu **distribuce dokumentů/řízení přístupu**. Tytéž snahy nalezneme i u zdrojů poskytujících sekundární informace. Abychom mohli v integračním systému uživatelům přinášet primární dokumenty a obsahy databází, musíme tato práva a požadavky zajistit.

Zatímco u papírových kopií dokumentů platí jasná pravidla daná fyzickou podstatou nosiče, zajištění práv u elektronických dokumentů je problematictější. Je třeba nasadit technologie, které dokáží řídit kopírování a modifikování dokumentů a kontrolovat přístup k jejich obsahu. Problém je třeba řešit ve dvou rovinách:

- na úrovni zabezpečení přístupu k dokumentu v integračním systému (ochrana **online**)
- na úrovni práce s dokumentem v rukách uživatele (ochrana **offline**)

Jinými slovy, ochrana online zabezpečuje **přístup k dokumentu**, ochrana offline pak zabezpečuje **přístup k obsahu** dokumentu. Nejprve je třeba zajistit, aby se k daným informačním zdrojům na integračním serveru dostali pouze ti uživatelé, kteří k tomu mají oprávnění. Když oprávněný uživatel dokument získá a stáhne si ho na svůj lokální počítač, je třeba dále zajistit, aby uživatel nemohl dokument neoprávněně kopírovat nebo ho dále distribuovat, přivlastnit si autorství, popř. aby po vypršení oprávnění nemohl s dokumentem dále pracovat. V mnoha případech je cílem distribuce dokumentu pouze transport k uživateli pro účely **jednorázového tisku**, po kterém by měl dokument **expirovat**.

Ochrana online spočívá hlavně v zajištění bezpečné autentikace (viz kapitola 2.3) a celkového zabezpečení integračního systému (viz kapitola 2.4), tedy serveru a připojených prvků. V praxi je bohužel pozornost často zaměřena pouze tímto směrem.

Ochranu offline, tedy ochranu samotných dokumentů po stažení z integračního serveru, kdy již nelze nijak ovlivnit další distribuci, zajišťují **technologie dokumentových formátů**. V řadě realizovaných řešení je tento aspekt zcela opomíjen a ochrana obsahu dokumentů je ignorována. Dokumenty jsou tak často dále distribuovány ilegálními cestami, využívány neoprávněnými osobami a nekontrolovaně modifikovány a kopírovány. V dokumentech je proto nutné pro zajištění práv a licencí zajistit trvalou informaci o nositeli autorských práv, řídit přístup k obsahu dokumentu, řídit práva k manipulaci s dokumentem, řídit čas expirace s následným znemožněním přístupu a zajistit kompaktní formátování dokumentu.

Řada technologií na ochranu autorských práv a licencí využívá šifrovacích algoritmů, které se dělí na symetrické a asymetrické.

**Symetrické šifrovací algoritmy** používají stejný klíč pro šifrování i pro dešifrování. Klíč musí zůstat utajen, aby se neoprávněné osoby nemohly dostat k šifrovanému obsahu. Klíč musí být často střídán a musí být dostatečně náhodný.

**Asymetrické šifrovací algoritmy** používají k šifrování a dešifrování vždy různé klíče, z nichž jeden je veřejný a druhý soukromý. Soukromý klíč je číselná hodnota chráněná heslem, umožňující autorovi podepisovat dokumenty. Veřejný klíč autora se přibaluje k podpisu a slouží k identifikaci a ověření autentičnosti původce/podpisu na straně příjemce. Veřejný klíč příjemce může autor použít k přidělení oprávnění pracovat s šifrovaným dokumentem. Veřejné klíče nemusí být utajovány, protože bez příslušných soukromých klíčů nemohou odhalit obsah dokumentů.

### 3.3.1 Označení původce dokumentu a vlastníka autorských práv

Nejběžnějšími technologiemi, které umožňují do dokumentů zapisovat a následně v nich identifikovat původce, jsou metadata, digitální podpisy a vodoznaky. Všechny tyto metody lze využít v integračním systému jako součást procesu publikování a distribuce.

**Metadata** jsou popisné údaje o obsahu a charakteru dokumentu a mohou být uložena přímo v datovém souboru nebo externě mimo soubor. Možnost zápisu metadat přímo do datových souborů se objevuje v nových verzích téměř všech moderních dokumentových formátů, a to jak textových, obrazových, zvukových tak i ve videoformátech. Metadata mohou kromě jiného obsahovat údaje o původci dokumentu, nositeli autorských práv, možnostech užití, roku vzniku atd. a lze je proto využít jako pomůcku pro ochranu autorských práv. Příkladem mohou být metadata ve formátu MS Word, PDF, JPG (metadata EXIF nebo IPTC), MP3 (metadata ID3) atd. Metadata jsou editovatelná, což samo o sobě nezajišťuje účinnou ochranu, protože je může uživatel přepsat nebo odstranit. Výjimkou jsou formáty schopné nastavit vnitřní práva pro práci se souborem a tak i znemožnit editace metadat, jako je např. PDF.

**Vodoznaky** u elektronických dokumentů jsou založeny na podobném principu jako vodoznaky na papíře. Jedná se o aplikaci grafického vzoru, obrázku nebo textu na pozadí nebo popředí textového nebo obrazového dokumentu, často s vysokou mírou průsvitnosti, aby vodoznak neznemožnil prohlížení dokumentu. Jejich účelem je označit a identifikovat dokument a vlastníka autorských práv. Při správně zvolené aplikaci jsou relativně těžko odstranitelné, a proto z hlediska uchování informací o autorských právech značně efektivní. Nevýhodou aplikace vodoznaků je změna původního obsahu, případně i snížená čitelnost. Vodoznaky lze do

dokumentů snadno vkládat v běžných textových/grafických editorech nebo ve speciálně k tomu vybavených systémech.

Pokročilejší a účinnější variantou vodoznaku je **digitální vodoznak**, jehož prostřednictvím je do obrazového, ale i zvukového nebo video dokumentu vložena informace o autorských právech v podobě okem neviditelného nebo uchem neslyšitelného šumu. Důležité je, že vložený digitální vodoznak nijak nenarušuje kvalitu obsahu daného dokumentu. K výhodám této technologie patří snadná detekce a relativně vysoká odolnost vodoznaku, který se neztrácí ani při výrazných modifikacích dokumentu, jako je komprese nebo převzorkování (odolnost závisí na použitých metodách vkládání). Různá aplikační řešení navíc umožňují prohledávat určené části internetu pomocí robotů a zjišťovat, nevyskytují-li se někde na cizích serverech naše dokumenty v rozporu s licenčními podmínkami. Možnosti kontroly nad nepovolenou distribucí a neoprávněným použitím dokumentů jsou proto poměrně vysoké. Vývojem technologií digitálního vodoznaku se zabývá řada IT firem, nejznámější z nich je Digimarc s technologií MyPictureMarc<sup>13</sup>, kterou podporuje i Adobe Photoshop. Firma Digimarc navíc nabízí i zajímavou technologii digitálního vodoznaku pro textové dokumenty na principu úpravy šířky mezer mezi slovy a řádkování.

**Digitální podpisy** představují asi nejelegantnější způsob označování a identifikace vlastníka autorských práv dokumentu. Technologie digitálního podpisu vznikly pro účely ověřování identity původce dokumentu, důvěryhodnosti informací a integrity dokumentu (možnosti sledování změn v dokumentu po jeho vzniku). Při použití digitálního podpisu se do dokumentu vloží jedinečný otisk se zašifrovanými číselnými údaji. Digitální podpis může informovat každého z uživatelů o subjektu, který daný dokument vytvořil, a umožnit tak kontakt s nositelem autorských práv nebo řešení případných sporů týkajících se autorství.<sup>24</sup> Digitální podpis vychází většinou z asynchronní koncepce tzv. soukromého a veřejného klíče, kterou můžeme dále rozdělit do dvou skupin. První skupina technologií vyžaduje existenci tzv. certifikační autority, tedy jakéhosi garantovaného vydavatele platných podpisů. Druhá skupina se obejde bez certifikační autority a je založena na tzv. přímé důvěře jednotlivých účastníků výměny dokumentů. Pro účely ochrany autorských práv jsou vhodnější technologie s účastí certifikační autority kvůli nezávislé garanci platnosti a pravosti podpisů, což může být vhodné při možných soudních sporech. Nejvíce rozšířené standardy pro digitální podpisy jsou DSS (Digital Signature Standard) používající algoritmus DSA (Digital Signature Algorithm) a PKCS (Public Key Cryptography Standards) používající mimo jiné nejpoužívanější šifrovací algoritmus RSA (Rivest, Shamir & Adelman). Hotová řešení pro podepisování dokumentů nabízejí např. firmy Adobe, Microsoft, Entrust, VeriSign a GeoTrust.

### 3.3.2 Řízení přístupu k obsahu dokumentu

Základem ochrany dokumentu před neoprávněným užitím je **šifrování jeho obsahu**. Slouží jak pro účely distribuce dokumentu, kdy není zajištěna zabezpečená komunikace a dokument se může dostat do rukou neoprávněných osob, tak pro účely práce s dokumentem offline. Úspěšnost šifrování je dána mírou **snížení šance prolomit ochranu**, tj. nalézt dešifrovací klíč. Šifrovacích mechanismů je velké množství a stále vznikají dokonalejší (vývoj složitějších algoritmů je umožněn zejm. nárůstem výkonu počítačů). Šifrování a následné dešifrování se provádí pomocí hesla nebo jiného kódu, kterým může být např. bezpečnostní certifikát, podle toho, zda-li je použitá šifrovací metoda symetrická nebo asymetrická. Nástroje a řešení na šifrování dokumentů nabízejí např. firmy Adobe, OpenFile Systems nebo Microsoft.

Dokument může být šifrován již na straně vzdáleného informačního zdroje nebo ho může šifrovat až integrační systém. O odpovědnosti za obsah v integračních systémech rozhoduje způsob distribuce dokumentů. Získává-li uživatel dokument přímo ze vzdáleného informačního zdroje (integrační systém uživatele pouze přesměřoval), nese odpovědnost za zabezpečení obsahu vzdálený zdroj. Pokud ale integrační systém vystupuje jako přímý distributor dokumentů, tj. uživatel získává dokumenty jeho prostřednictvím, musí integrační systém zajistit ochranu vlastními prostředky.

### 3.3.3 Řízení práv k manipulaci s dokumentem

Jakmile uživatel získá elektronický dokument do svých rukou, může s ním provádět řadu operací, které mohou být v rozporu s licenční politikou. Integrační systém by proto měl před vlastní distribucí dokument ošetřit, aby k porušování licencí nedocházelo. Je to jediný způsob,

jak získat důvěru vydavatelů, producentů a provozovatelů v bezpečnou digitální distribuci dokumentů a navazující služby.

Klíčovými předměty ochrany jsou:

- **tisk** – uživatel může dokument vytisknout na tiskárně nebo mohou jiné aplikace využívat tiskový výstup k importu obsahu (aplikace využívající tiskové ovladače)
- **kopírování a vyjímání obsahu dokumentů** – uživatel může přes schránku OS kopírovat obsah do jiných aplikací (např. textového editoru) nebo mohou k obsahu přistupovat jiné aplikace a vytěžovat ho
- **změny v obsahu dokumentu** (editace) – uživatel může měnit texty, obrázky či jiné objekty v dokumentu, čímž může změnit nejenom podobu ale smysl dokumentu
- **změny v uspořádání dokumentu** – uživatel může přidávat, vyjímát či otáčet stránky nebo měnit jejich pořadí, čímž porušuje původní podobu originálního dokumentu

Výše uvedené operace s dokumentem je možné kontrolovat nastavením uživatelských práv, pokud to formát dokumentu umožňuje. Jediná operace, kterou nelze spolehlivě ošetřit, je sejmutí dokumentu na obrazovce jako snímek (screenshot), protože zobrazení dokumentu na obrazovce je předpokladem čtení a obrazovku lze snímat prostředky OS. Restrikce by měly být vždy založeny na šifrování dokumentu, aby byly účinné i v aplikacích jiných výrobců software, které nemusí tyto ochranné technologie respektovat. Stejně jako u řízeného přístupu lze šifrovat heslem (synchronní šifrování) nebo certifikátem (asynchronní šifrování).

Některé operace je však často výhodné pro dokumentové formáty uživatelům nebo jiným aplikacím umožnit. Patří sem např. zajištění přístupu k metadatům dokumentu i v případě, že je dokument šifrován, možnost vkládat uživatelských poznámek apod. Tyto vlastnosti podporuje např. formát PDF zmiňovaný v kapitole 3.3.6.

### 3.3.4 Řízení expirace platnosti a znehodnocení dokumentu

Přístup k obsahu dokumentu lze limitovat také z časového hlediska. Uživatel může získat oprávnění využívat obsah dokumentu pouze do určitého okamžiku. Tento okamžik označujeme jako expirace. Podpora řízení expirace dokumentu nachází uplatnění v celé řadě případů. Pro účely služeb integrovaného systému jsou to zejm.:

- **přístup k obsahu dokumentu pro účely pořízení tištěné kopie** – expirace je nastavena na krátkou dobu nebo pro jednorázové otevření, aby měl uživatel možnost dokument vytisknout (elektronická kopie slouží pouze pro účely distribuce)
- **realizace digitálních výpůjček** – jedná se o elektronickou podobu klasických výpůjček např. na 30 dní, poté je přístup k obsahu uživateli odepřen
- **vynucení aktualizace dokumentu online** – u dokumentů, jejichž obsah má význam pouze v případě, že je aktuální, lze omezením doby platnosti donutit uživatele pořizovat si pravidelně aktuální kopie

Technologie řízení expirace dokumentů jsou většinou založeny na **šifrování** obsahu a na otisku **časového razítka** (time stamp), což je časový záznam digitálně podepsaný důvěryhodnou třetí stranou, který vyjadřuje okamžik expirace. Na podobném principu je založena ochrana software s časově platnými licencemi nebo ochrana dočasných práv na využití určitých, většinou placených služeb.

S expirací je spjat problém možného přenastavení systémového času na koncovém zařízení, kdy uživatel po vypršení oprávnění předstírá, že okamžik expirace ještě nenastal. Tento problém se pokouší řešit řada technologií, nejčastěji pomocí ověření skutečného času u tzv. autority časových razítek, která garantuje pravdivost časového údaje. Tyto metody však bohužel vyžadují účast vzdálených systémů, tedy předpokládají, že bude uživatel online a že ověření časového údaje povolí.

Účinné řízení expirace lze u dokumentů dosáhnout pomocí DRM systémů (Digital Rights Management), mezi které patří např. produkt OpenFile Publisher od firmy OpenFile Systems, PageRecall od firmy Authentica, WMDRM od firmy Microsoft nebo řešení LiveCycle od firmy Adobe.

### 3.3.5 Zajištění kompaktního formátování dokumentu

Snaha o zajištění kompaktního a neměnného formátování textu a grafiky v dokumentech je vždy výsledkem kompromisu. Na jedné straně se snažíme vyhovět licenčním podmínkám poskytovatelů a uživatelům zajistit dokument, který se za všech okolností zobrazí stejně a čitelně, na druhé straně musíme vyhovět různým potřebám uživatelů (poruchy zraku, obliba jiných stylů) a koncových zařízení (různá úroveň zobrazovacích schopností). Zatímco v letech, kdy se objevovaly první projekty pro zpřístupňování dokumentů přes internet, byly nejběžnějšími dokumentovými formáty HTML a prostý text, dnes jsou dokumenty šířeny mnoha různými formáty. Velkého rozšíření se dočkal formát PDF, stále častěji se objevují textové formáty odvozené z XML (např. RSS). Mění se také spektrum používaných koncových zařízení - stále více uživatelů používá k přístupu na web jinou techniku, než jsou počítače (mobily, PDA).

Při hledání vhodného výstupního formátu z IS pro integraci informačních zdrojů je vhodné posoudit následující otázky:

- **Je třeba zabránit změnám formátování z licenčních důvodů?** Tento požadavek bývá spojen s požadavkem na neměnnost obsahu a věrohodnost dokumentu. V takovém případě je vhodné použít kompaktní formáty s pevným stránkováním a rozložením, jako je např. PDF nebo obrazové formáty. Ty zároveň zajišťují, že je formátování a rozložení textu stejné při zobrazení i při tisku. Alternativou mohou být formáty podporující různé „zámky“ zakazující změny v dokumentu (např. Microsoft Word a funkce Zámek), ty ale většinou neřeší problém změny velikosti zobrazovacího okna a následné přeformátování.
- **Jedná se o dokumenty převážně textové?** U textových dokumentů není problém v dynamickém přeformátování textu podle šířky zobrazovacího zařízení, proto pro tento případ dobře poslouží formáty průtokového typu, např. HTML nebo RTF.
- **Obsahují dokumentu velké množství grafiky nebo komplikované rozložení?** Obsahuje-li dokument velké množství tabulek, obtékání grafických objektů apod., může změna šířky zobrazovacího okna text zdeformovat a narušit rozložení s negativním vlivem na čitelnost, srozumitelnost i význam. Zde se nabízí více řešení. Kromě použití kompaktních formátů, jako je PDF, lze objekty zajistit proti změnám prostřednictvím CSS, definování absolutních nebo minimálních rozměrů objektů apod. Ale vzhledem k tomu, že integrační systém využívá velké množství různorodých informačních zdrojů s rozmanitým spektrem formátů, je vhodné zvolit jednoduchý způsob, který půjde aplikovat univerzálně bez nutnosti vytvářet adaptéry pro konkrétní případy.
- **Budou dokumenty dodávány i na mobilní zařízení?** Mobilní zařízení jako PDA nebo mobilní telefony s OS mají relativně malou zobrazovací plochu. Pro tyto zařízení je vhodné formátování textu upravit, eliminovat grafiku, případně grafické objekty zobrazovat pouze na vyžádání. Široce podporovaným formátem pro tato zařízení je kromě použití prostého textu a HTML formát Mobi PRC s možností zabezpečení, který podporují např. firmy Nokia, Sony Ericsson, Siemens a Motorola.

Pokud může integrační systém ovlivnit výstupní formát dokumentů, lze pro textové dokumenty, které není třeba chránit, obecně doporučit formát HTML. Pro dokumenty s komplikovaným rozložením nebo pro dokumenty vyžadující ochranu je obecně nejvýhodnější formát PDF a pro mobilní zařízení formát Mobi PRC.

### 3.3.6 Formát PDF pro komplexní řízení ochrany dokumentu

K dispozici je dnes několik široce podporovaných dokumentových formátů, které podporují všechny výše zmíněné vlastnosti dokumentů a pro které jsou nabízeny ucelené sady aplikačních nástrojů pro editaci, indexaci a prohlížení dokumentů v těchto formátech. Jedná se např. o formát LIT od firmy Microsoft nebo formát Mobi PRC pro mobilní zařízení od firmy MobilPocket. Řada výrobců také používá různé modifikace otevřeného formátu OEB (Open E-book), který rozšiřují o vlastní metody šifrování a zabezpečení. Tyto technologie jsou poměrně dynamicky rozvíjeny díky vytvářejícímu se trhu s elektronickými knihami a časopisy, kde je zabezpečení obsahu dokumentu klíčovým předpokladem komerčně úspěšného prodeje titulů.

Nejflexibilnější a nejvíce podporovanou technologií pro zajištění ochrany autorských práv a vyhovění různým licenčním a bezpečnostním požadavkům je však bezesporu dokumentový formát PDF (Portable Document Format) od firmy Adobe. Formát PDF ve verzi 1.6 a technologie

s ním spojené mají několik klíčových výhod, které z nich dělají favorita pro bezpečnou distribuci dokumentů v rámci integračního systému:

- formát PDF je hned po HTML nejpoužívanějším dokumentovým formátem na webu
- prohlížeč PDF je nezávislý na OS a instalovaných písmech
- prohlížeč PDF je bezplatný, bez požadavku na investice do SW u příjemce dokumentu
- PDF zachovává vzhled a integritu originálních dokumentů
- texty mohou být do PDF vloženy jako texty, křivky nebo bitmapy
- do PDF formátu je možno vkládat bitmapové obrázky i vektorové objekty různých typů
- metadata formátu PDF jsou uchovávána v XML
- PDF umožňuje dynamické přeformátování textu pro účely přenosných zařízení
- export do PDF je podporován pro většinu jednoduchých i profesionálních kancelářských a grafických editorů, včetně DTP a CAD systémů
- PDF podporuje plnotextové prohledávání, vkládání poznámek, záložek a datových polí
- PDF umožňuje vkládání jakýchkoli elektronických příloh, včetně multimediálních
- PDF umožňuje zabezpečení dokumentu proti neoprávněnému prohlížení, tisku, kopírování obsahu, včetně nastavení času expirace, po kterém již není možné dokument použít
- zabezpečení je postaveno na šifrování obsahu dokumentu a/nebo jeho příloh
- dokumenty v PDF lze digitálně podepsat a tím označit původce dokumentu
- dokumenty v PDF lze dodatečně opatřit vodoznaky, pozadím, záhlavím a zápatím
- PDF plní i archivní funkci se zárukou, že půjde dokument otevřít i po letech nezávisle na hardware a OS (formátová podmnožina PDF/A)
- PDF plní také funkci elektronické obálky pro bezpečnou distribuci objektů

Firma Adobe nabízí pro formát PDF ucelené aplikační řešení v podobě dokumentového serveru Adobe LiveCycle Document Server<sup>2</sup>. Tento systém zajišťuje tvorbu dokumentů (např. skenování, převodem z jiných formátů, vkládáním údajů z databází nebo XML do šablon), správu a zabezpečení a řízenou distribuci dokumentů jak v elektronické, tak tištěné podobě (PDF, tisk, fax, e-mail, web, bezdrátový přenos). Dokumenty může vzájemně sdružovat, šifrovat, nastavovat úroveň přístupu a digitálně podepisovat. Jako modulární systém s otevřeným rozhraním ho lze bezešvě integrovat do ucelených integračních řešení. Představuje proto ideální stavební kámen při budování IS pro integraci informačních zdrojů.

PDF řešení lze použít jako univerzální výstup z integračního systému pro jakýkoli typ dokumentů i strukturovaných záznamů, protože do tohoto formátu lze snadno převádět jakýkoli tiskový výstup nebo formátovat data dynamicky z libovolných datových zdrojů. Díky možnosti zabezpečení PDF tak můžeme bezpečně distribuovat plné texty, referáty, záznamy i grafické objekty.

### 3.4 TECHNOLOGIE VZDÁLENÉHO PŘÍSTUPU

Provozovatelé často zpřístupňují své informační zdroje na základě licenčních ujednání s důrazem na dodržování autorských a vydavatelských práv. Při integraci informačních zdrojů je proto třeba uspokojivě vyřešit problém, jak těmto právům a licenčním podmínkám dostát, ale jak zároveň neznemožnit uživatelům s informačními zdroji **pracovat kdykoli a odkudkoli**.

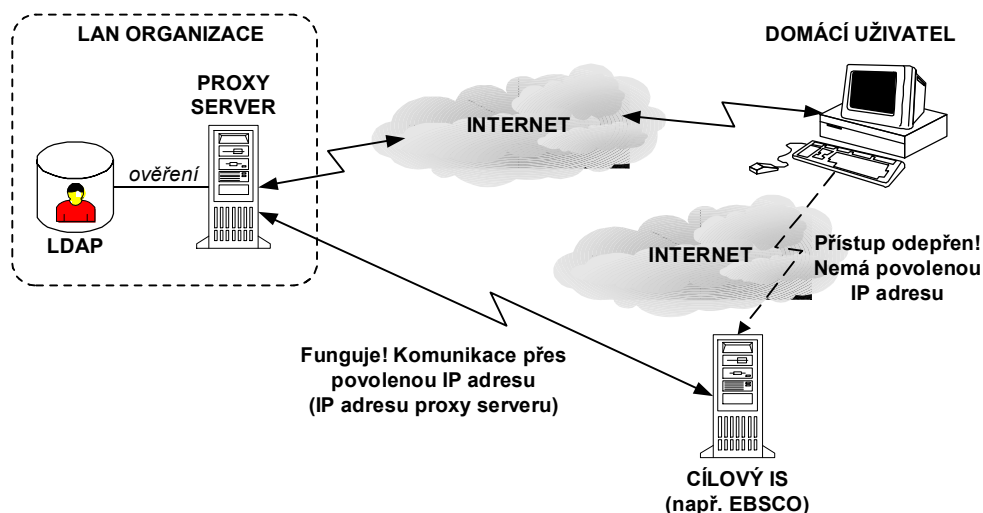
Informační zdroje, zvláště pokud jsou komerční, se většinou snaží zamezit neoprávněným přístupům k obsahu a vyžadují různé druhy autentikace. Nejčastěji se v praxi setkáváme s autentikací na základě ověření uživatelského jména a hesla, s autentikací na základě IP adresy počítače nebo s kombinací obojího (více informací o autentikaci – viz kapitola 2.3). Uživatelská jména a hesla pro přístup k informačním zdrojům nesmějí instituce svým uživatelům sdělovat, IP adresy jsou vázány na rozsahy dané instituce. Výsledkem je, že uživatelé nemohou přistupovat k obsahu informačních zdrojů mimo prostory instituce, např. ze zaměstnání nebo z domova. Řešení tohoto problému nabízejí technologie vzdáleného přístupu.

Technologie vzdáleného přístupu musí být pro uživatele snadno použitelná a zároveň bezpečná. Uživatelé nejsou většinou ochotni nebo schopni na svá zařízení nic instalovat, navíc koncových zařízení existuje celá řada a jsou platformově velmi rozmanitá. Nabízí se však několik typů vzdáleného přístupu, které se v současné době díky velké poptávce po těchto technologiích rychle vyvíjejí, zdokonalují a vůči koncovému uživateli zjednodušují. Pro účely integrace

informačních zdrojů jsou nejvyužívanější technologie vzdáleného přístupu: proxy server, VPN, terminálový přístup a Shibboleth.

### 3.4.1 Proxy server

Proxy server je počítač nebo specializovaný software, který umožňuje uživatelům nepřímé připojení k jinému serveru. V případě integrace informačních zdrojů hraje proxy server důležitou úlohu v tom, že funguje jako prostředník mezi uživatelem a cílovým informačním zdrojem, překládá jeho požadavky a vůči cílovému informačnímu zdroji vystupuje jako klient. Přijatou odpověď posílá následně zpět uživateli.



Obr. 18 – Vzdálený přístup k informačnímu zdroji přes proxy server

Vzdálený informační zdroj, ke kterému uživatel přistupuje prostřednictvím proxy, tak komunikuje pouze s proxy serverem a netuší, jaký uživatel ho stínově využívá – všechny počítače, které proxy server využívají, se jeví jako jeden počítač s IP adresou proxy serveru. Z hlediska autentikace pak stačí, aby byla všechna oprávnění na straně vzdáleného informačního zdroje nastavena pouze pro daný proxy server, např. na základě jeho IP adresy či loginu a hesla. Proxy server však musí v takovém případě garantovat, že na něj budou přistupovat pouze oprávnění uživatelé, např. čtenáři konkrétní knihovny. Uživatelé se tak musí pro využití proxy serveru autentikovat, typicky uživatelským jménem a heslem, nejlépe šifrovaně. K tomu účelu se většinou využívá adresářových služeb přístupných přes LDAP, aby nemusela pro potřeby proxy vznikat samostatná databáze uživatelů.

Nasazením proxy serveru je v integračním systému dosaženo:

- **ochrany soukromí koncových uživatelů**
- **zvýšení výkonu komunikace** (pokud proxy server funguje zároveň jako cache a odpovědi si ukládá do vyrovnávací paměti)
- **zvýšení bezpečnosti** (uživatelé nemusí znát login a heslo ke vzdáleným zdrojům, navíc komunikace může být šifrována)
- **tvorby centralizovaných statistik přístupů** ke všem zdrojům
- **vzdálený přístup k informačním zdrojům odkudkoli**, protože již nehraje úlohu IP adresa uživatele

Funkci proxy serveru může plnit integrační systém již v principu na základě své architektury, např. pokud funguje jako paralelní vyhledávač, který požadavky uživatelů distribuuje na vzdálené servery a vystupuje tak jako prostředník v procesu vyhledávání.

### 3.4.2 VPN

VPN (Virtual Private Network) je technologie, která typicky slouží pro připojení vzdáleného počítače do lokální sítě organizace. I když se může počítač nacházet ve fyzicky nezávislé síti



kdekoli na světě, může prostřednictvím VPN komunikovat se síťovými prostředky lokální sítě, jako by byl v lokální síti fyzicky přítomen. K navázání VPN spojení je třeba VPN server, který má přístup na internet i do lokální sítě organizace, a VPN klient, který se přes internet připojí k VPN serveru a prostřednictvím něho pak do lokální sítě. VPN server vystupuje jako síťová brána, která s klientem vytvoří zabezpečený síťový tunel, v rámci něhož pak probíhá požadovaná síťová komunikace. Podle konkrétně použité technologie jsou tak přenášená data většinou šifrována a zapouzdřena, často silnými šifrovacími algoritmy (např. DES, Triple DES nebo AES). Uživatelé pomocí VPN získají IP adresu instituce, která VPN provozuje, čímž získají přístup ke všem zdrojům, které jsou zprovozněny v dané lokální síti.

Klienti VPN jsou dnes již součástí všech operačních systémů, navíc existuje řada VPN klientů určených pro konkrétní síťové technologie, např. firmy Cisco. Připojení k VPN serveru je však nutné na straně klienta nakonfigurovat, což vyžaduje určitou technickou zdatnost. Proto se připojení VPN využívá zejm. na akademické půdě, méně již ve veřejných knihovnách.

### 3.4.3 Terminálový přístup

Terminálový přístup zajišťuje vzdálené přihlášení ke vzdálenému počítači, typicky serveru. Uživatel pomocí speciálního programu, který spouští na svém počítači, pracuje na vzdáleném počítači stejným způsobem, jako kdyby u něho fyzicky seděl – na vstup jsou přenášeny příkazy z klávesnice a myši, na výstupu se vracejí generované obrazovky. Terminálový přístup může být zajištěn pouze v textovém režimu (např. telnet) nebo v grafickém režimu (např. Vzdálená plocha v MS Windows). Velkou výhodou je, že vzdálený počítač může pracovat pod zcela jiným operačním systémem než počítač uživatele. Uživatel může využívat různé aplikace a zdroje, které jsou na vzdáleném počítači k dispozici – proto se terminálový přístup také označuje jako sdílení pracovní plochy (Desktop Sharing) nebo VNC (Virtual Network Computing). Komunikace při terminálovém přístupu bývá navíc šifrovaná, takže práce je bezpečná.

Nasazení technologie terminálového přístupu je vhodná u integračních projektů s malou uživatelskou základnou, protože je značně náročná na výkon terminálového serveru, zvláště pokud pracuje v grafickém režimu. Výhodou tohoto řešení je garance programového vybavení uživatele, protože uživatelské prostředí je realizováno na serveru.

### 3.4.4 Shibboleth

Technologie Shibboleth je představitelem tzv. **distribuované autentikace**. Zajišťuje přístup k zabezpečeným lokálním i vzdáleným informačním zdrojům na základě důvěryhodného ověření identity uživatele, navíc mezi zúčastněnými IS vytváří prostředí **SSO**. Technologie je navržena a dále rozvíjena konsorciem Internet2.

Prostředí Shibboleth vyžaduje spolupráci poskytovatelů obsahu (informačních zdrojů) a informačních systémů zajišťujících ověřování identity uživatelů. Jinými slovy, aby mohla být technologie Shibboleth nasazena, musí ji podporovat jak příslušné informační zdroje, tak instituce, které zajišťují ověřování identity svých uživatelů. Lze však také nasadit tzv. Shibboleth agregátory, které technologii Shibboleth podporují a které na základě IP adresy nebo loginu a hesla poskytují autentikaci pro ty informační zdroje, které technologii Shibboleth nepodporují.<sup>35</sup> Takovým agregátorem je např. produkt EZproxy od firmy Useful Utilities.

Systém pro správu identity uživatelů (Identity Management System) slouží k poskytování služeb potřebných pro identifikaci osob působících v určitém systému a pro řízení přístupu identifikovaných uživatelů k těm komponentám systému, ke kterým je přístup nějakou formou omezený. Jedná se tedy o komplex služeb, který poskytuje na základě definovaných politik bezpečné a automatizované řešení správních úkonů efektivně řešících přístup uživatelů k omezeným a mnohdy citlivým informačním zdrojům.<sup>16</sup>

Protože je toto prostředí založeno na vzájemné důvěře, jsou zřizovány tzv. federace poskytovatelů obsahu a poskytovatelů identit. Základem technologie Shibboleth je služba Where Are You From (WAYF), která zajišťuje přesměrování autentikačních požadavků od poskytovatelů obsahu na příslušného poskytovatele ověření identity uživatele. Uživatel, který chce přistoupit k zabezpečenému informačnímu zdroji, je přesměrován na službu WAYF, kde si zvolí svého

poskytovatele identity a u něho se autentikuje. Informační zdroj následně získá výsledek autentikace, případně minimální sadu atributů nutných pro autorizaci služby informačního zdroje. Poté může oprávněný uživatel začít se zdrojem pracovat.

Technologicky je Shibboleth založen na otevřených standardech SAML a LDAP, čímž je dána jeho snadná implementovatelnost ve stávajících prostředích IS. Softwarové nástroje Shibboleth jsou k dispozici jako open source pro všechny běžné operační systémy, čímž je zajištěn předpoklad postupného rozšíření mezi provozovatele informačních zdrojů. V současné době již Shibboleth podporují provozovatelé Ebsco, ProQuest, Elsevier, Blackwell, JSTOR, OCLC, Thomson Gale, CSA, Ovid Technologies a mnoho dalších.

### 3.4.5 Technologie SSO

Single sign-on (SSO) představuje skupinu speciálních technologií, které umožňují uživateli využívat **jeden autentikační kód** (běžně uživatelské jméno a heslo) pro přístup do více nezávislých IS. Typickým představitelem těchto technologií jsou centralizované adresářové služby typu LDAP nebo Active Directory, které slouží k ukládání autentikačních a autorizačních politik jednotlivých IS při společné databázi uživatelů. V kombinaci s distribuovanou autentikací může SSO umožnit uživateli **autentikovat se pouze jednou** a následně automaticky získávat přístup do více nezávislých IS. Tuto rozšířenou formu SSO reprezentují např. protokoly SAML či již zmíněný Shibboleth - systém SSO přijme autentikační token jiného IS, ověří ho a následně umožní uživateli využít požadovanou službu či informační zdroj, aniž by se musel uživatel znovu přihlašovat.

V kontextu integrace informačních zdrojů je SSO předpokladem integrace často různorodých autentikačních mechanismů, kterými jsou informační zdroje zabezpečeny.

Kromě zvýšení komfortu práce koncových uživatelů, kteří si díky SSO nemusí pamatovat desítky přihlašovacích jmen a hesel, přináší tyto technologie mnoho výhod i pro samotný integrační systém. Jedná se zejména o **sjednocení autentikačních a autorizačních politik** v rámci celého integračního systému, zlepšení možností pro **provádění bezpečnostních auditů**, provozování **jediného úložiště identit** uživatelů pro všechny komponenty integračního systému a **snížení zátěže na podporu** koncových uživatelů v otázkách zapomenutých hesel, nesprávně zapisovaných hesel či nepřístupných účtů.

K dispozici je řada řešení, které je možné bezplatně použít v integračních projektech. Kromě již zmíněných technologií LDAP, Shibboleth a SAML se jedná zejm. o populární open source řešení Central Authentication Service (CAS) vyvinuté na univerzitě v Yale, projekt CoSign podporovaný sdružením National Science Foundation Middleware Initiative (NMI) či široce oblíbená technologie Kerberos.

## 3.5 TECHNOLOGIE PRO VYTVÁŘENÍ OBSAHU INFORMAČNÍCH ZDROJŮ

Instituce, které provozují IS pro integraci informačních zdrojů, často sami vytvářejí informační objekty, které publikují v podobě elektronických dokumentů nebo informačních databází. Stávají se tak sami producenty, případně provozovateli dílčích informačních zdrojů, které mohou být objektem zájmu jejich vlastního nebo jiného integračního systému. Z toho důvodu by měly volit takové publikační technologie, které umožní snadnou a bezešvou integraci. Podívejme se v krátkosti na současné možnosti publikování elektronických informací.

### 3.5.1 Digitální repozitáře

Zamýšlí-li instituce budovat sbírku statických elektronických dokumentů, je vhodným řešením digitální repozitář. Digitální repozitář slouží jako **organizované úložiště digitálních objektů** a nemá žádné ambice řídit jejich vznik či následnou editaci. S objekty zachází jako s uzavřenými entitami, může k nim však připojovat popisná **metadata** vytvářené buď ručně nebo na základě extrakce údajů z příslušných dokumentových formátů (např. z XMP metadat formátu PDF nebo z EXIF metadat formátu JPEG). Součástí metadat by měl být vždy jednoznačný identifikátor objektu, volitelně pak různé formy věcného popisu, např. klasifikační schémata či abstrakta.

Digitální repozitář by měl v první řadě umět vysoce organizovaně ukládat a evidovat všechny objekty a plnotextově i strukturovaně v nich vyhledávat. Má tak mnoho společných rysů se systémy DMS (Document Management System), které se využívají ve firemní oblasti jako jeden z nástrojů podnikových intranetů. Digitální repozitář by měl být také vybaven rozhraním OAI-PMH pro sklizení metadat jinými informačními systémy, případně dalšími formami distribuce.

**Tvorba a editace digitálních objektů je typicky realizována mimo digitální repozitář,** často jako výsledek digitalizace klasických fondů nebo publikační činnosti dané komunity (např. zveřejňování závěrečných prací na vysokých školách).

### 3.5.2 Publikační systémy

Pokud chce instituce spojit fázi **tvorby** elektronických dokumentů s fází **publikování** v jednom systému, je vhodné použít systém pro správu obsahu. Tyto systémy se obecně označují zkratkou CMS (Content Management System) nebo v případě zúžení na webový obsah jako WCM (Web Content Management).

Součástí CMS bývá redakční modul, který umožňuje provádět rychlé změny obsahu dokumentů (většinou HTML dokumentů) bez znalosti programování a formátovacího jazyka. Zároveň často podporuje postupy redakčního schvalování, čímž umožňuje týmovou spolupráci více autorů při tvorbě obsahu. Vytvářený obsah je automaticky plnotextově indexován pro účely vyhledávání, zároveň je možné dokumenty opatřovat popisnými metadaty. Dokumenty jsou začleněny do přísně organizované struktury složek a podsložek, což přináší další možnosti v navigaci k obsahu. Přestože standardním rozhraním CMS bývá webové rozhraní, k dispozici bývají i další možnosti přístupu k obsahu, např. RSS kanály či rozhraní OAI-PMH. Další volitelnou součástí CMS mohou být nástroje umožňující interakci s uživateli, např. uživatelská fóra, ankety, možnost vkládat komentáře k článkům atd. CMS se pak přibližuje architektuře portálů.

Systémy CMS jsou vhodné k tvorbě dynamických webů, kde se předpokládá spolupráce týmu autorů, k publikování článků ve formě elektronických časopisů apod.

### 3.5.3 Databáze

K publikování **strukturovaných dat**, reprezentujících ať již primární či sekundární informace, jsou vhodným řešením aplikace využívající nějaký systém pro řízení bází dat (DBMS). Klasickým příkladem takových řešení jsou automatizované knihovní systémy, ve kterých se budují katalogy knihoven či jiné bibliografické databáze, nebo plnotextové databáze článků. Typicky tyto aplikace obsahují editační i prezentační část, takže může instituce v jednom systému informační objekty vytvářet i v nich vyhledávat. Architektura těchto aplikací většinou přísně odděluje aplikační a databázovou vrstvu, což má příznivý vliv na výkon, flexibilitu i bezpečnost.

Při výběru databázových systémů by měl být kladen důraz na interoperabilitu všech forem komunikace, směrem do systému i směrem do okolí. Tu typicky zajišťují standardizované jazyky vyšší úrovně pro manipulaci a definici dat, např. SQL nebo QBE. Klíčovými hledisky při výběru databázového systému by měla být robustnost a zotavitelnost při chybách bez ztráty dat. Moderní systémy DBMS proto často podporují správu jednotlivých datových transakcí a technologie zajišťující integritu dat.

Nejvyužívanějšími DBMS jsou v současné době z komerční oblasti technologie Oracle Database 10g, Sybase Adaptive Server Enterprise 15, Microsoft SQL Server 2005, IBM DB2 9 či IBM Informix 10. Jsou však k dispozici i vysoce kvalitní DBMS šířené nekomerčně, např. MySQL 5, PostgreSQL 8 nebo SQLite 3.

## 3.6 SHRNUTÍ

**Při projektování IS můžeme vybírat z bohaté nabídky informačních technologií, které nám dnešní trh nabízí. Nejsou to pouze komerční technologie, čím dál častěji se objevují kvalitní technologie šířené zdarma jako GPL či freeware. Pokud se rozhodneme pro vývoj vlastních proprietárních řešení, měli bychom k tomu mít vážný důvod – ztrácíme tak totiž řadu výhod otevřených standardů a ověřených technologií. Použité technologie by měly být dostatečně otevřené a flexibilní, abychom je mohli**

**v případě potřeby nahrazovat jinými. Ideální je nasazení modulární architektury složené ze samostatných komponent integrovaných mezi sebou do funkčního celku. Při výběru technologie musíme posuzovat celkové náklady na vlastnictví, nejen pořizovací cenu. Na celkových nákladech se totiž podílí i různé udržovací poplatky, mzdové náklady na odborníky a náklady na podporu, cena za hardware apod. po celou dobu životního cyklu.**

**V IS pro integraci informačních zdrojů hrají nejdůležitější úlohu technologie pro vyhledávání a komunikaci se vzdálenými informačními zdroji, technologie na zajištění ochrany autorských práv a licencí, technologie vzdáleného přístupu a technologie pro vytváření obsahu. Protože informační technologie se rychle a dynamicky mění, je nutné stále sledovat trendy a vývojové směry, zejm. z hlediska standardizace.**

## 4 INTEGRACE INFORMAČNÍCH ZDROJŮ A ŘÍZENÍ JAKOSTI

V předešlých kapitolách jsme již mnohokrát narazili na skutečnost, že se informační zdroje, které do našeho IS integrujeme, vyznačují různou technickou vyspělostí, spolehlivostí, stabilitou, kvalitou a důvěryhodností obsažených informací, úplností apod. Uvnitř integračního systému probíhá řada automatizovaných i manuálních procesů, které mohou být různě řízeny a nastavovány, čímž lze ovlivňovat celkové chování IS. Pokud chceme na výstupu IS zajistit kvalitu služeb odrážející požadavky uživatelů, musíme pečlivě sledovat, vyhodnocovat a ovlivňovat každý příslušný proces či komponentu. Takové řídicí činnosti obecně označujeme jako **systém řízení jakosti** (quality management systém, zkráceně QMS). Systém řízení jakosti musí být nasazen ve všech etapách projektování IS, který popisuje kapitola 1.6.

Řešení pro integraci informačních zdrojů musí být technicky i organizačně stabilní, efektivní a spolehlivé. Toho lze při řízení jakosti dosáhnout:

- vytvořením dosažitelného dlouhodobého obchodního modelu pro zajištění financování
- změnou koncepcí organizace, pokud je to pro dosažení cílů třeba
- vybudováním vztahů s provozovateli informačních zdrojů a jejich dlouhodobé řízení
- definováním a sledováním úrovní služeb
- vytvořením a sledováním kvality poskytovaných informací a služeb

Obecně můžeme systém řízení jakosti vyvinout sami, nicméně z hlediska ověřených postupů a spolupráce s okolím je vhodné zavést normalizovaný systém řízení jakosti označovaný jako ISO 9000. Jedná se o ucelenou skupinu norem (ISO 9000-4), které jsou celosvětově uznávány a jsou spojeny s možností oficiální certifikace, která slouží zákazníkům jako důkaz o spolehlivosti a důvěryhodnosti IS a celé organizace. Je vhodné se jimi inspirovat i v případě, že je nechceme do organizace přímo zavádět. V dnešní praxi se využívá revize ISO 9000 z roku 2000.

Normy ISO 9000 předpokládají, že stabilní prosperity dosáhne organizace zaměřená na zákazníka. Jakostí se zde rozumí schopnost splnit požadavky zákazníka. Zákazník je spokojen přímo úměrně tomu, jak dalece byly splněny jeho požadavky. **Jakostní služba je taková, se kterou je zákazník spokojen.** Díky takto obecně formulovaným požadavkům je systém možno aplikovat v téměř jakékoli organizaci. Stačí si dosadit za službu naši nabídku informačních služeb nabízenou prostřednictvím IS pro integraci informačních zdrojů. Jsou-li naše služby kvalitní, je splněn základní předpoklad vysokého využívání a oblíbenosti našich služeb. Vysoká míra využití zajistí životaschopnost celého integračního projektu. Organizace může díky neustálému zlepšování uspokojovat potřeby zákazníků i dlouho po tom, co byl IS nasazen.<sup>45</sup>

Jednotlivé funkce, které monitorujeme, mohou být upravovány ještě dalšími normami, aby vyhověly funkčním požadavkům. K takové oblasti patří zejm. garance bezpečnosti IS, kterou upravuje norma ISO/IEC 15408 (Kritéria pro hodnocení bezpečnosti IT). Nasazení bezpečnostních funkcí do IS je předpokladem zajištění důvěryhodnosti, integrity a dostupnosti.<sup>14</sup>

Všimněme si nyní specifických požadavků normy ISO 9000 z hlediska integrace informačních zdrojů podle členění samotné definice této normy<sup>44</sup>:

### 4.1 SYSTÉM ŘÍZENÍ JAKOSTI

Při realizaci větších integračních projektů je nezbytné zavést automatizovaný systém řízení jakosti pro **identifikaci všech procesů a vazeb** v IS a jejich následné sledování. Nejdůležitější procesy jsou ty, které mají přímou vazbu na koncového uživatele. Do integračního systému jsou typicky zapojeny samostatně provozované, heterogenní komponenty (informační zdroje a externí služby), které nemůžeme přímo technicky ani obsahově ovlivňovat. Důraz na monitoring a benchmarking těchto komponent je pro zajišťování jakosti celého IS zcela klíčový.

Je třeba permanentně sledovat kvalitu připojení a chování všech začleněných informačních zdrojů, zejm. zpracováním a analýzou logů, reportů a statistik generovaných informačním systémem, důsledným evidováním licencí a kontrolou oprávnění přístupu k informačním zdrojům. Systém řízení jakosti v této oblasti musí proto obsahovat prvky **ERM systémů**.

## 4.2 ODPOVĚDNOST VEDENÍ

Vedení projektu pro integraci informačních zdrojů je přímo spjato se strategickým plánováním a nasazováním politiky jakosti. Plánovací proces projektu musí periodicky přezkoumávat efektivitu systému řízení kvality a řešit případné nedostatky. Centrem pozornosti celého systému musí být koncový uživatel, a proto by měly být všechny procesy směřovány k naplnění uživatelských požadavků a informačních potřeb. Koncový uživatel musí být vnímán jako zákazník, cílem je jeho **spokojenost**.

Vedení zároveň pověřuje pravomocemi a odpovědností další členy týmu, případně externí spolupracovníky. Klíčové je rozdělení odpovědnosti za technický stav a obsah mezi vlastní IS (resp. jeho lokální část) a mezi distribuované komponenty (vzdálené informační zdroje a externí služby). Je také důležité uživatele přesně informovat, kdo zaručuje **důvěryhodnost** každého dílčího informačního objektu. To může být při slučování a agregaci informací velký problém, nicméně je nezbytné údaje o odpovědnosti a původci informace v celém procesu integraci zachovávat, aby se ve výsledné podobě prezentace vždy objevily. Na straně výstupů je také třeba přesně definovat, nakolik se odpovědnost přenáší na vlastní koncové uživatele, zejm. v oblasti autorských práv.

Předpokladem fungujícího řízení je funkční **komunikační systém** uvnitř organizace/týmu, aby mohli na konkrétní situace pružně reagovat všichni kompetentní členové.

## 4.3 ŘÍZENÍ ZDROJŮ

Projekt musí efektivně využívat **finanční a lidské zdroje**. Kvalitní obchodní model a personální politika je nutným předpokladem úspěšnosti projektu. Nezbytné je také definovat požadavky na **infrastrukturu**, zejm. v oblasti ICT (kvalitní síťové připojení, software a hardware) a samotných informačních zdrojů. Požadavky musí být realizovatelné a udržitelné po celou dobu trvání projektu, a to ve shodě s požadavky na služby koncovým uživatelům.

Důraz je kladen na potřebnou kvalifikaci, odbornost a zaškolení členů týmu, aby byly minimalizovány všechny **negativní dopady lidského faktoru** na fungování IS. Odbornou způsobilost nestačí zajistit pouze na počátku projektu nebo náboru nového člena týmu, ale je třeba zajistit její průběžné udržování, případně její zlepšování. Významnou podporu výtěžnosti lidských zdrojů také představuje řádně určené a řízené pracovní prostředí.

## 4.4 REALIZACE SLUŽEB

Vedení projektu plánuje a rozvíjí procesy potřebné pro realizaci integračního projektu. Základem těchto plánů musí být **sběr a přezkoumávání požadavků koncových uživatelů**. Sběr požadavků musí být realizován řízenou komunikací, aby byla umožněna zpětná vazba a možnost reakce. Podrobnou analýzou, selekcí a následným tříděním specifických požadavků definujeme návrhy na další vývoj IS a příslušných funkcí. Maximální pozornost musíme věnovat externím dodavatelům, kteří mají **přímý vliv na uživatele** (ať již samotní provozovatelé připojovaných informačních zdrojů nebo dodavatelé softwarových komponent). Je vhodné se snažit o propojení našeho systému řízení kvality s obdobnými systémy našich dodavatelů, resp. dodavatele můžeme vybírat mimo jiné také podle schopnosti vyhovět normám ISO 9000. Požadavky na jakost externích služeb je nutné zapracovat do dodavatelských smluv a průběžně jejich plnění kontrolovat.

Musíme přesně určit, jaké procesy přímo ovlivňují kvalitu nabízených služeb, a ty potom průběžně **sledovat a měřit** jejich parametry. K tomu je třeba definovat, jaké monitorovací a měřicí prostředky použijeme.

Typickými parametry, které bychom měli u IS pro integraci informačních zdrojů sledovat, jsou zejm. míra dostupnosti služeb (vyjádřená v % v konkrétním časovém období), stabilita IS, důvěryhodnost informací, rychlost dodávání primárních informací, kontrola jakosti manuálních služeb, bezpečnost IS s návazností na autentikační a autorizační mechanismy, prokazování

shody s platnými oborovými standardy, možnosti přístupu uživatelů do IS (možnosti vzdáleného přístupu), akvizice informačních zdrojů, míra překrývání obsažených informačních zdrojů atd.

#### 4.5 MĚŘENÍ ANALÝZY A ZLEPŠOVÁNÍ

Musíme dbát na to, aby byl nasazený systém řízení jakosti skutečně v projektu uplatňován. K tomu slouží **interní audit**, který bychom měli pravidelně opakovat. Postup auditu systému managementu jakosti popisuje norma ISO 19011. Pro audit dílčích komponent či procesů IS jsou k dispozici specializované normy, jako např. ISACA 2003, IS auditing procedure – firewalls.

Pro monitoring a měření procesů je třeba definovat metriku a postup vyhodnocování. Odhalené nedostatky musíme okamžitě eliminovat pomocí preventivních opatření dříve, než se problémy projeví u koncového uživatele. V případě vážných závad je třeba mít připraveny havarijní plány.

Neustálým sběrem a analýzou vybraných údajů získáváme informace o míře spokojenosti koncových uživatelů a o shodě služeb s jejich informačními potřebami. V průběhu provozování integračního řešení musíme shromažďovat zkušenosti a systém řízení jakosti musíme neustále zlepšovat.

#### 4.6 CERTIFIKACE

Organizace provozující IS pro integraci informačních zdrojů může požádat o certifikaci systému řízení jakosti. Akreditovaná certifikační organizace formou auditu zkoumá, nakolik se námi zavedený systém řízení jakosti shoduje s požadavky norem ISO 9000, a v případě shody vydává osvědčení.

Některé organizace pochopily řízení jakosti pouze jako snahu získat příslušný certifikát a ne jako trvalé zlepšování managementu. Takový přístup je chybný a nemůže vést k dlouhodobému udržení spokojenosti uživatelů.

## 5 MODELOVÝ PŘÍKLAD INTEGRACE INFORMAČNÍCH ZDROJŮ

V této kapitole si ukážeme, jakým způsobem by mohla být řešena integrace informačních zdrojů na modelovém příkladu společného informačního portálu knihoven v Hradci Králové, který bude na jednom místě integrovat informační zdroje a služby. Cílem tohoto modelu je ukázat možné technologické řešení, nejsou zde proto obsaženy žádné pasáže o časovém harmonogramu realizace, vstupní analýzy ani postup pro výběr dodavatelů. Při sestavování tohoto modelu byly využity zkušenosti a metodika z reálného projektu e-Government na Novém Zélandu a Jednotné informační brány Národní knihovny ČR.

### 5.1 ÚVODNÍ ZPRÁVA – SOUČASNÁ SITUACE

Hradec Králové je krajské město, kde žije přibližně 100 000 obyvatel. Ve městě působí 6 významných knihoven\*, které jsou zaměřeny na různé typy uživatelů, nicméně každá z těchto knihoven nabízí služby i pro širší veřejnost. Jsou to tyto knihovny:

- Studijní a vědecká knihovna v Hradci Králové
- Knihovna města Hradec Králové
- Knihovna Univerzity Hradec Králové
- Knihovna Lékařské fakulty v Hradci Králové Univerzity Karlovy v Praze
- Knihovna Farmaceutické fakulty v Hradci Králové Univerzity Karlovy v Praze
- Vědecká knihovna Fakulty vojenského zdravotnictví Univerzity obrany

V současné době mohou obyvatelé Hradce Králové využívat informačního potenciálu a služeb knihoven, které jsou ve městě k dispozici, pouze separovaně a s omezenými možnostmi. Uživatelé jsou nuceni pracovat odděleně s dílčími informačními systémy každé z knihoven, do každého systému se registrovat a pro realizaci mnoha služeb knihovnu osobně navštívit.

Jednotlivé hradecké knihovny nabízejí různé formy služeb v různé kvalitě, kooperace mezi nimi je velmi pasivní. Při akvizici informačních zdrojů nijak nespolupracují, navzájem ani neuznávají registrované uživatele.

Cílem projektu je vybudovat společný **Informační portál královéhradeckých knihoven**, který umožní jednotlivým knihovnám konsolidovat na jednom místě své informační zdroje a služby. Uživatel tak prostřednictvím portálu získá přístup k celému informačnímu portfoliu šesti největších knihoven města, a to kdykoli a odkudkoli. Služby integrované na jednom místě, dostupné online na dálku by měly eliminovat potřebu uživatelů fyzicky navštěvovat knihovnu, čímž se služby otevrou i pro osoby se specifickými potřebami či pro uživatele ze vzdálenějších koutů regionu, kteří těžko využívají zdroje lokálně. Portál by se měl stát jedinou zastávkou na cestě při získávání všeho, co uživatel potřebuje. Bude uživatele provázet od okamžiku rozpoznávání, jaké služby a nástroje uživatel potřebuje, přes získání přístupu k těmto službám a jejich využití. Portál by měl vždy nabízet alternativy v získávání informací, pokud nějaká služba nemůže být z různých příčin využita. Cílem je spokojený uživatel, který nikdy neodejde „s prázdnou“.

### 5.2 VIZE

Informační portál královéhradeckých knihoven se stane hlavním poskytovatelem knihovních informací a služeb regionu prostřednictvím internetu a dalších komunikačních kanálů. Stane se upřednostňovaným způsobem přístupu uživatelů ke knihovním službám, bude zohledňovat specifické potřeby a zvyky obyvatel města a regionu. Stane se synonymem královéhradeckých knihoven.

---

\* Kromě toho ještě ve městě působí několik neveřejných nebo speciálních knihoven, které jsou omezeny pouze pro potřeby svých zaměstnanců nebo badatelů – ty nebudeme v integračním projektu uvažovat.



### 5.3 CÍL

Cílem Informačního portálu královéhradeckých knihoven je integrovat informační zdroje a navazující služby královéhradeckých knihoven na jediné místo, které bude přístupné online pro všechny zájemce s rozlišením jejich specifických potřeb. Portál bude nabízet pohodlný, rychlý, finančně nenáročný a bezbariérový přístup. Portál bude přísně orientován na koncové uživatele. Díky modernímu a efektivnímu způsobu doručování služeb bude portál propagovat práci knihoven, které se tak těsněji dostanou do povědomí občanů města a regionu.

Informační portál královéhradeckých knihoven bude moci využít kdokoli s přístupem na web, určen je však zejména pro:

- obyvatele města Hradec Králové
- obyvatele královéhradeckého regionu (menších měst a vesnic)
- studenty, pedagogické a vědecké pracovníky
- lidem, kteří se zajímají o královéhradecký region

Uživatelé portálu budou kategorizováni podle věku, profesního zaměření, jazyka, míry počítačové gramotnosti a dalších charakteristik, aby pro ně mohly být segmentovány služby a nabídky napříč celým portálem. Uživatelé však budou moci výchozí nastavení funkcí kdykoli individuálně změnit a zobrazovat pouze komponenty, které jsou pro ně přínosné.

Služby portálu budou mezi sebou propojeny, aby uživatelé při procesu získávání a zpracování informací nemuseli opakovaně zadávat vstupní údaje. Vyhledávání informací bude pouze prvním krokem v procesu uspokojování informačních potřeb uživatelů.

S ohledem na různé úrovně počítačové gramotnosti bude každá samostatná část nebo funkce portálu vybavena srozumitelnou nápovědou, průvodcem nebo možností využití online podpory.

### 5.4 ÚKOLY

Pro dosažení cíle je třeba definovat strategické úkoly, které se budou plnit v průběhu budování portálu. Z nich bude odvozen a sestaven podrobný harmonogram.

Cíle	Metrika	Termín
Portál se stane preferovaným způsobem při využívání informací a služeb hradeckých knihoven.	Cíle bude dosaženo, pokud více než polovina uživatelů daných knihoven bude využívat služeb prostřednictvím portálu.	do 1,5 roku
Portál bude v sobě integrovat veškeré uživatelské služby, které hradecké knihovny nabízejí online.	Cíle bude dosaženo, když všechny online služby zúčastněných knihoven budou přístupné prostřednictvím portálu.	do 1 roku
Do portálu budou zapojeny i ty služby, které nejsou při zahájení projektu k dispozici online.	Cíle bude dosaženo, pokud portál umožní využívání offline služeb elektronickou cestou prostřednictvím portálu.	do 2 let
Spustit portál jako plně samoobslužný.	Cíle bude dosaženo, jakmile bude veškerá podpora využívání služeb k dispozici online.	do 4 let
Celkové pojetí nabízených služeb bude orientováno na uživatele a bude odstíněn přístup jednotlivých knihoven.	Cíle bude dosaženo, když více než 75% uživatelů potvrdí, že portál splňuje jejich požadavky nezávisle na zvyklostech jednotlivých knihoven.	do 6 měsíců
Portál nabídne uživatelům prostor pro vzájemnou spolupráci v rámci odborných nebo občanských komunit.	Cíle bude dosaženo, pokud portál umožní vzájemnou komunikaci a sdílení mezi jednotlivými uživateli portálu, minimálně na úrovni zasílání offline zpráv a sdílení složek.	do 3 let

## 5.5 POŽADAVKY NA FUNKCE

- Portál musí umět integrovat informační zdroje, které nabízejí různorodé komunikační rozhraní a formáty dat, a umožňovat jejich prohledávání v jednotném rozhraní.
- Vyhledávací postupy by měly podporovat jak hledání konkrétních, předem známých dokumentů, tak provádění rešerší, kdy uživatel předem neví, jaké dokumenty hledá.
- Portál musí poskytovat služby zúčastněných knihoven a napříč mezi zúčastněnými knihovnami. Z pohledu uživatele by měly být jednotlivé knihovny jako provozovatelé dílčích služeb odstíněny.
- Portál by měl nabízet průvodce pro získávání primárních informací podle různých scénářů a situací.
- Uživatelé musí mít možnost si portál personalizovat a přednastavit chování všech modulů. Budou si moci vytvářet vlastní seznamy a skupiny zdrojů, ukládat rešeršní dotazy i vyhledané záznamy. Měl by nabízet komunikační kanály pro všechna zařízení, která uživatelé běžně používají (PC, notebooky, PDA, mobilní telefony).
- Přístup bude umožněn anonymním návštěvníkům i registrovaným uživatelům. Registrovaní uživatelé budou mít k dispozici širší nabídku služeb s možností personalizace a nabídku informačních zdrojů obohacenou o komerční zdroje podle oprávnění.
- Uživatelé budou mít k dispozici uživatelskou schránku, se kterou získají diskový prostor jako úložiště informací, komunikační nástroje pro offline i online komunikaci (e-mail, instant messaging), možnost sdílet části prostoru s ostatními uživateli, využívat informační avíza a adresné šíření informací pro bezobslužné informování o změnách v nabídce informačního portfolia.
- Portál bude otevřený i pro další poskytovatele informací, které se budou chtít později do portálu připojit a nabízet své informační služby.
- Všechny součásti portálu a jednotlivé informační kanály bude možno zabezpečit v souladu s dikcí autorského zákona a zákona na ochranu osobních informací.
- Portál bude přístupný 24 hodin denně, 7 dní v týdnu. V rozsahu 16 hodin denně po celý týden bude zajištěna technická podpora.
- Všechny uživatelské procesy v portálu budou statisticky sledovány, aby bylo možné podle situace měnit nastavení portálu a nabídku informačních zdrojů a služeb.
- Použité technologie budou přísně splňovat mezinárodně platné webové a knihovnické standardy a budou pokud možno využívat již existující infrastrukturu zúčastněných knihoven.

## 5.6 PŘIDANÁ HODNOTA PROJEKTU

Informační portál královéhradeckých knihoven přinese řadu výhod pro uživatele i zúčastněné knihovny. Tyto výhody představují přidanou hodnotu ke službám, které nyní nabízejí zúčastněné knihovny samostatně.

### 5.6.1 Přidaná hodnota pro uživatele

- **Rovnocenný přístup** k informačním službám knihoven bez ohledu na věk, bydliště, vzdělání, profesi, jazyk a další faktory s ohledem na speciální potřeby uživatelů
- Výrazně **rychlejší přístup** k informačním službám knihoven
- **Snadnější a přehlednější přístup** k nabídce služeb díky sofistikovaně navrženému uživatelskému rozhraní portálu
- **Jediné uživatelské rozhraní** pro práci s mnoha různorodými lokálními i vzdálenými informačními zdroji
- Možnost přistupovat k řadě informací a služeb **anonymně**
- Informační zdroje a služby využitelné vzdáleně mimo prostory knihoven **kdykoli a odkudkoli**
- Všechny služby dosažitelné na jednom místě prostřednictvím **jednoho z nabízených komunikačních kanálů**

### 5.6.2 Přidaná hodnota pro zúčastněné knihovny

- Všechny zúčastněné knihovny budou hrát v portálu **rovnocennou roli**, spolupráce bude založená na **vzájemné důvěře** a smyslu pro společný projekt.
- Knihovny budou své služby nabízet **jednotným způsobem** při dodržení společných standardů, čímž uživatelé získají celou nabídku služeb konzistentně a uspořádaně. Zároveň však bude zachována nezávislost jednotlivých knihoven.
- Zpřístupněné informační zdroje budou nabízeny **v souladu s autorskými a licenčními právy**.
- Díky jednotnému a přehlednému přístupu k nabídce služeb **vzroste využití služeb**.
- Díky integraci vznikne možnost **objektivně porovnávat služby** jednotlivých knihoven.
- Jednotlivé knihovny budou mít k dispozici nástroje, které by si kvůli finanční a technické náročnosti nemohly samy dovolit. **Sdílení** finančních, personálních a technologických zdrojů bude efektivní pro všechny účastníky projektu.
- Virtuální propojení dílčích fondů knihoven umožní **koordinovanou akvizici**, čímž dojde k efektivnějšímu využívání finančních prostředků na budování fondů.
- Centralizovaným poskytováním služeb dojde ke **snížení zátěže** na jednotlivé zúčastněné knihovny.
- Portál překoná mnoho komunikačních bariér, čímž dojde ke **zlepšení vztahů** mezi uživateli a knihovnami. Knihovny budou moci **aktivněji nabízet a propagovat svoje služby**.

### 5.7 KRITICKÉ FAKTORY

- **Plnění závazků zúčastněných knihoven** – knihovny musí splnit své závazky v projektu, zejm. spuštěním svých služeb online
- **Kvalita popisu a zpřístupnění služeb** – dostupné služby musí být snadno vyhledatelné a dosažitelné
- **Přednostní poskytování služeb prostřednictvím portálu** – portál musí pro knihovny představovat prioritní způsob zpřístupnění služeb
- **Ekonomická udržitelnost projektu** – projekt musí být zaštitěn promyšleným obchodním modelem, který bude zajišťovat stabilní finanční podporu
- **Propagace** – portál musí být dostatečně propagován, aby si uživatelé uvědomili jeho přínos a navykli si ho využívat
- **Schopnost spolupráce s jinými projekty** – portál musí být schopen systémově spolupracovat s okolím a vytvářet širší integrační celky
- **Průběžné informování o rozšiřování a zlepšování funkcí portálu** – uživatelé musí pravidelně dostávat informace o novinkách a plánovaných vylepšování portálu, aby věděli, že se portál vyvíjí a je zajištěna jeho budoucnost

### 5.8 ŘEŠENÍ

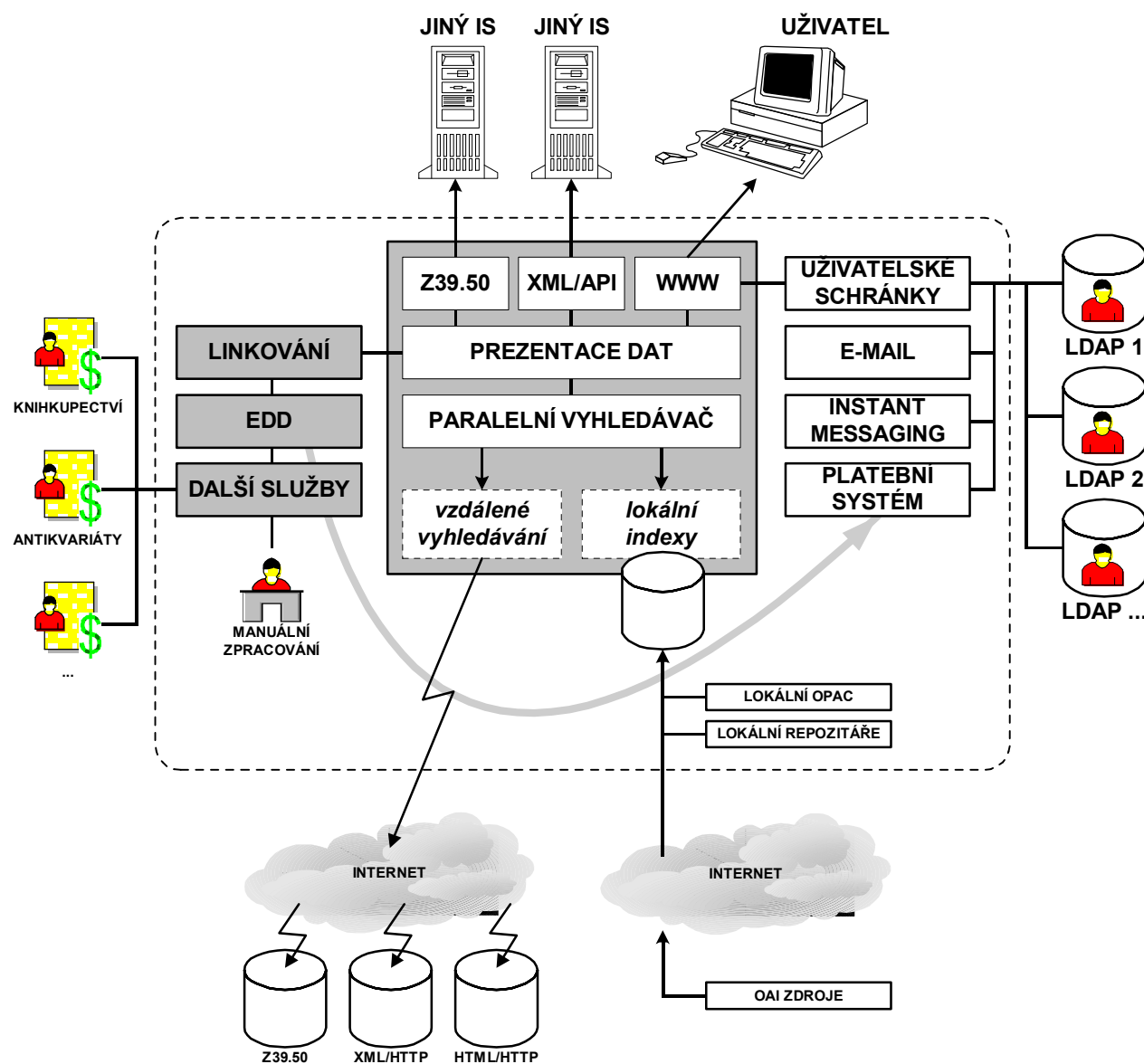
Jednotlivé informační zdroje budou do portálu připojeny prostřednictvím metavyhledávače. Metavyhledávač zajistí možnost paralelního vyhledávání jediným dotazem, a to vyhledávání jak lokálních, tak vzdálených zdrojů. V zájmu co nejširší podpory různých způsobů připojení budou k dispozici adaptéry Z39.50, SRU/SRW, obecné XML/HTTP, HTML/HTTP a úložiště pro sklizení přes protokol OAI-PMH. Metavyhledávač musí umět provádět deduplikaci a slučování vyhledaných informačních objektů do jedné prezentační skupiny výsledků.

Všechny připojené informační zdroje i služby budou podrobně popsány ve speciální databázi, která umožní uživatelům snadno a přesně nalézt a použít potřebné nástroje. Při výběru zdrojů a služeb budou preferovány ty, které nejlépe a nejrychleji zajistí přístup k primárním informačním objektům. Služby, které neprobíhají automaticky, budou na pozadí distribuovány k vyřízení jednotlivým zúčastněným knihovnám. Vstupy a výstupy manuálně prováděných služeb budou do portálu připojeny elektronicky. Uživatel se nikdy neocitne ve slepé uličce, vždy bude veden dál, dokud nebudou vyčerpány všechny možnosti získávání, které se nabízejí.

Nedílnou součástí řešení musí být závazná společná metodika pro popis a technické nastavení jednotlivých informačních zdrojů, které knihovny provozují lokálně, tj. jejich vlastní katalogy, databáze a repozitáře.

Vyhledávání bude možné na všech analytických úrovních (např. v databázích, časopisech i člancích), ve výsledcích budou tyto úrovně přehledně rozlišeny. Výsledné množiny informačních objektů bude možné řadit, a to i podle více kritérií současně, a třídit do fasetů podle vybraných polí.

Vazby na primární objekty budou realizovány buď přímo z vyhledaných výsledků nebo pomocí linkovacího serveru, který uživatele přesměruje do jiných zdrojů dat. Pokud nebude k dispozici elektronická podoba primárního objektu, bude mít uživatel možnost využít služby EDD nebo jiných doručovacích služeb, linkovací server zajistí podle potřeby přesměrování požadavku na komerční knihkupectví, antikvariáty či vydavatelství. K dispozici bude i paleta manuálně prováděných doplňujících služeb, zejm. odborných rešerší, reprografických služeb apod. V nabídce bude také poloautomatizovaný systém pro zadávání referenčních dotazů. Eliminovány budou čistě bibliografické databáze, které bez vazby na primární objekty nemají pro koncové uživatele žádný význam.



Obr. 19 – Architektura modelového IS

Zúčastněné knihovny si zachovají své databáze uživatelů, nedojde ke sloučení procesu registrace uživatelů. Knihovny však budou při poskytování služeb vzájemně uznávat registrované uživatele jiné zúčastněné instituce. Portál bude schopen autentikovat uživatele ve všech dílčích databázích uživatelů pomocí protokolu LDAP. Pro registrované uživatele bude k dispozici uživatelská schránka, která bude sloužit jako úložiště záznamů, rešerší, SDI, primárních informačních objektů apod. K uloženým objektům budou uživatelé moci připojovat své vlastní poznámky, podle kterých bude možno vyhledávat. Úložiště si budou moci rozčlenit do složek a tyto složky budou moci mezi sebou sdílet. Na uživatelskou schránku bude navázán bezpečný platební systém pro realizaci mikroplateb za nákup zpoplatněných informačních objektů. Uživatelé budou mít také možnost využít služby elektronické pošty a online komunikace (instant messaging).

Portál bude podporovat několik typů uživatelských rozhraní, aby s ním mohli uživatelé pracovat bez ohledu na typ zařízení a technické vybavení. Základní způsob přístupu bude zajišťovat WWW rozhraní, které na základě rozpoznání klienta zoptimalizuje výstup pro daný prohlížeč a zařízení. Dalšími typy rozhraní budou WAP a e-mail/SMS/RSS (pro rozesílání informací, avíz, zpráv SDI, nových přírůstků apod.). Bude zachována možnost telefonických dotazů, aby mohli systém využívat i uživatelé s omezenými schopnostmi práce s IT. K celému obsahu bude umožněn i vzdálený přístup, aby mohli uživatelé pracovat odkudkoli, nejen z prostor knihoven. Vzdálený přístup bude realizován na základě autentikace a šifrované komunikace.

Ve všech modulech budou logovány probíhající procesy a události, následně budou statisticky zpracovávány. Statistiky z logů spolu s podklady technické podpory a reakcí uživatelů budou složité jako hlavní zdroj dat pro zpětnou vazbu.

## 6 BEST PRACTICES

Jako „best practices“ se označují tzv. nejlepší postupy k řešení dané problematiky. Jsou to metodické návody, které jsou sice podepřeny **teorií**, ale vycházejí převážně z **praxe**. Sumarizují zkušenosti, které se nejlépe osvědčily v předchozích podobně zaměřených projektech, a tyto zkušenosti se snaží zobecňovat, aby byly lehko aplikovatelné a modifikovatelné na další modely. Cílem je vyvarovat se chyb, kterých se již někdo před námi dopustil, a naopak vybrat řešení, které se ukázalo jako účinné. Organizace tak mohou díky best practices dosahovat ve svých projektech dobrých výsledků, přestože nemají k dispozici specialisty na všechny relevantní oblasti. Využívání metod best practices představuje schopnost organizace učit se od ostatních.

Je však třeba zdůraznit, že best practices nejsou definitivní, neměnné a trvale nejlepší. Vždy je možné najít ještě lepší postup, který posune vývoj a zkušenosti dopředu. Význam best practices je v tom, že popisují **nejosvědčenější** a nejlepší známá řešení v dané době a **v daném kontextu**. Best practices tak při hledání vlastních řešení vystupují jako náměty. Nejsou to závazné předpisy, a proto poskytují dostatečný prostor k vylepšování, hledání a objevování. Pojetí nejlepších postupů v sobě odráží princip **nepřetržitého zlepšování kvality** v souladu s očekáváním zákazníků, který známe z oblasti řízení výroby – japonsky označovaný jako „kaisen“.

Myšlenku best practices poprvé zmínil Frederick Taylor na počátku 20. století: „mezi různými metodami a nástroji používanými v dílčích fázích každé aktivity existuje vždy jedna metoda a jeden nástroj, které jsou rychlejší a lepší než všechny ostatní“. Tato myšlenka začala být označována jako „jeden nejlepší způsob“.18 Od té doby, vlivem technického a technologického rozvoje, význam nejlepších postupů stoupal až do dnešní podoby. Dnes můžeme v každém odvětví lidské činnosti, které je typické rychlými změnami technologických nástrojů a nutností pohotově reagovat na změny, nalézt odborné aktivity definující nejlepší postupy a řešení. Je to předpoklad standardizace, spolupráce, udržení konkurenceschopnosti a technické úrovně organizace. Best practices směřují každý projekt k nepřetržitému zlepšování, prodloužení životního cyklu a ideálně k jeho trvale udržitelnému rozvoji.

Tato kapitola přináší souhrn nejlepších postupů a zkušeností, které vznikly srovnáváním (benchmarkingem) tří velkých integračních projektů v průběhu jejich příprav a následných realizací. Jedná se o český celonárodní projekt **Jednotná informační brána** (JIB)<sup>43</sup>, který reprezentuje distribuovaný model integrace informačních zdrojů, o slovenský projekt **KIS3G**<sup>20</sup>, který se snaží vytvořit centralizovaný model AKS, a novozélandský vládní projekt **e-Government**<sup>15</sup>, který se soustředí na integraci informačních služeb veřejné správy. Analýzou a hodnocením těchto případových studií bylo stanoveno, jaké postupy, opatření a řešení se osvědčily a jaké naopak selhaly nebo vedly k nežádoucím výsledkům.

Celý proces benchmarkingu probíhal legálním způsobem, všechny použité informace jsou z právního hlediska veřejné. Mnoho z nich jsem však získal na základě důvěrné osobní komunikace, proto v nich byly obsaženy i negativní zkušenosti a postupy, které se neosvědčily a které nejsou záměrně zveřejňovány. Díky nim jsem mohl analýzu zkušeností rozšířit a výsledky zpřesnit z hlediska prevence nesprávných kroků. V získané informace mám díky kvalitním partnerům potřebnou důvěru. **Příprava** následující sady nejlepších postupů pro integraci informačních zdrojů probíhala v těchto krocích:

- definování cílů a jejich rozsah (základní představa o integračním systému a co má splňovat)
- sestavení seznamu dílčích oblastí měření a srovnávání (benchmarking)
- výběr projektů pro srovnávání
- sběr informací (zpracování projektové dokumentace, analýza implementovaných vlastností, rozbor systémových statistik, osobní návštěvy projektových týmů)
- analýza dat a tvorba znalostí
- výběr myšlenek a postupů a jejich zobecnění
- sestavení nejlepších postupů

Všechny zde uvedené nejlepší postupy byly vypracovány s ohledem na tyto **základní cíle**:

- úspěšné zprovoznění toho, co bylo v projektu naplánováno
- snížení nákladů

- zajištění a zlepšení kvality a rychlosti poskytovaných služeb
- maximální využití pořízených technologií
- zvýšení prestiže organizace
- lepší kontrola a vliv vedení na IS

Jednotlivé nejlepší postupy jsou přehledně rozděleny do oddílů, které odpovídají členění předchozích kapitol této práce, aby bylo možné sledovat vztah teoretických a praktických aspektů dané problematiky.

## 6.1 PLÁNOVÁNÍ A ŘÍZENÍ INTEGRAČNÍHO PROJEKTU

### 6.1.1 Zahájení projektu integrace informačních zdrojů

#### Popis problematiky

Instituce, které zahajují projekt integrace informačních zdrojů, musí mít zcela jasnou představu, čeho chtějí projektem dosáhnout. Nejčastějšími důvody bývá snaha vytvořit jednotné informační prostředí pro koncové uživatele, snaha propagovat služby knihovny v podobě portálu či snaha nabídnout uživatelům vzdálený přístup ke službám knihovny. Ambice mohou být veliké, ale konkrétní cíl projektu musí být reálný. Je lépe začít s méně ambiciózním projektem, která bude knihovna schopna v krátkém čase zrealizovat, a v další fázi na výsledky navázat dalšími projekty.

#### Praktické rady

- Projekt musí mít konkrétní a jasné cíle. Tyto cíle musí být zdokumentovány.
- Cíle musí odrážet přidanou hodnotu, kterou instituce realizací projektu získá.
- Cíle musí být realizovatelné vzhledem ke skutečným možnostem.
- Z cílů musí být odvozeny konkrétní úkoly, jejichž plnění je pravidelně měřeno.

### 6.1.2 Zvážení rizik projektu

#### Popis problematiky

Každý projekt by měl vznikat s cílem úspěšné realizace a nasazení. Avšak ani kvalitně sestavený projekt nemůže eliminovat všechna rizika, která mohou způsobit vážné komplikace nebo i nezdar celého projektu. Celkový úspěch projektu pak závisí na schopnosti s riziky dopředu počítat a umět na ně pružně reagovat účinnými změnami v architektuře IS, organizaci atd. Součástí všech fází projektu by proto měla být analýza rizik.

#### Praktické rady

- Rizika je třeba hledat nejen uvnitř projektu, ale i v jeho okolí. Jedná se zejména o možné změny legislativy (např. autorského zákona), změny technologií externích informačních zdrojů, ukončení vývoje použitého software apod. Architektura i pojetí celého IS musí umět reagovat na všechny změny v okolí.
- Největší pozornost je třeba věnovat rizikům s nejtěžším možným dopadem. Jedná se zvláště o:
  - zajištění financování projektu
  - odhad nákladů a doby trvání jednotlivých etap projektu
  - garance cen nakupovaných komponent, kvalita dodavatelských smluv (záruky)
  - možný odchod významného člena z týmu (závislosti na personálním zajištění)
  - příchod jiných trendů a změny priorit dané organizace
  - neustále se měnící požadavky
  - nedostatečné testování v jednotlivých fázích projektu
  - nedostatečné vazby mezi moduly IS
  - nedostatečná zpětná vazba od koncových uživatelů
- Mnoho rizik do projektu vnáší lidský faktor. Členy týmu je třeba dobře motivovat, průběžně je vzdělávat a manuálně prováděné činnosti kontrolovat. Mezi rizika způsobená lidským faktorem patří:
  - špatná pracovní morálka, malé zaujetí pro projekt
  - opomíjení a opoždování při plnění úkolů

- špatné porozumění úkolům
- chybné odhady
- nedostatečné dovednosti a znalosti pro plnění úkolů

### 6.1.3 Kritické aspekty projektování integračního systému

#### Popis problematiky

Přestože je každý projekt pro integraci informačních zdrojů více či méně specifický, můžeme definovat základní principy, které jsou svým charakterem a vzájemnou poslouplostí v praxi ověřeny. Zobecněný postup budování IS může sloužit jako námět a předloha při sestavování konkrétních integračních projektů.

#### Praktické rady

- V předprojektovém stádiu je třeba sestavit soupis zdrojů, které chceme ve střednědobém horizontu integrovat. Soupis musí vycházet ze skutečných potřeb uživatelů. Snahy integrovat „všechno“ jsou nerealizovatelné.
- Je třeba stanovit, jakým způsobem budou zdroje získány a jak budou využívány. Akvizice může být řešena buď přímým nákupem, na základě časově omezené licence (např. databáze), formou nákupu konkrétních článků na vyžádání, nákupem záznamů nebo slovníkových hesel na vyžádání, využitím externí služby EDD, spoluprací s jinou knihovnou atd. Je třeba počítat s tím, že rozpočty na akvizici se mohou rok od roku lišit.
- U každého uvažovaného informačního zdroje je nutné zjistit, jaká rozhraní nabízí. Z toho jsou pak odvozeny požadavky na technické řešení komunikačních adaptérů IS.
- Mnoho informačních zdrojů je možné získat od více producentů a provozovatelů, často za jiných finančních, technických a licenčních podmínek. Je třeba vybrat varianty, které jsou nejvýhodnější z hlediska zamýšlené integrace.
- Nezbytné je přesně zmapovat informační požadavky uživatelů nejen z hlediska informačních zdrojů, ale i z hlediska navazujících služeb, preferované formy služeb atd.
- Při projektování je prospěšné, seznámíme-li se s podobnými, již realizovanými projekty. Zkušenosti z nich můžeme získat v podobě metodických příruček, závěrečných zpráv, na stránkách příslušných projektů na internetu nebo formou konzultací se členy projektu.
- Klíčové je rozhodnutí, bude-li pro účely projektu vyvíjen proprietární software, bude-li software nakoupen jako hotové řešení nebo bude-li využíván jako služba. Jednotlivé možnosti se mohou kombinovat v rámci dílčích modulů IS. Výběr software má následně přímý vliv na výběr hardware.
- Informační zdroje je vhodné zapojovat do IS postupně. Pro pilotní spuštění postačí, když bude zapojeno jen několik základních zdrojů, aby bylo možné provést důkladné otestování.
- Každý informační zdroj by měl v IS procházet životním cyklem, jehož fáze budou po celou dobu evidovány (např. akvizice, popis, technická adaptace, ladění pro paralelní vyhledávání, napojení na navazující služby). Postup zapojování a zpracování zdroje by měl odpovídat fázím životního cyklu.
- Každá jednotlivá služba IS by měla být opatřena mechanismem zpětné vazby, abychom průběžně sledovali, jak uživatelé s IS pracují. Odhalené problémy a slabé stránky musí průběžně ovlivňovat další vývoj IS.

### 6.1.4 Personální zajištění projektu

#### Popis problematiky

Řada projektů doplatila na to, že vedení podcenilo personální zajištění a důsledkem toho došlo k opoždění jednotlivých etap, neuspokojivě řešeným dílčím úkolům nebo dokonce k úplnému selhání plnění. Lidé mají sklon k tomu, že podceňují pracovní a hlavně časovou náročnost úkolů, a považují za zbytečné vytvářet další pracovní pozice v projektu.

#### Praktické rady

- Ke každému dílčímu úkolu z projektového plánu je třeba přidělit potřebný počet pracovníků. Měla by být minimalizována závislost na konkrétních lidech, zvláště u dlouhodobých projektů.



- Podle předpokládané doby trvání a náročnosti jednotlivých úkolů je třeba zvážit, do jakých pracovněprávních vztahů s jednotlivými pracovníky vstoupíme. V některých případech může být výhodnější najmout si k provedení úkolu firmu.
- Pracovníky je třeba před plněním úkolu dostatečně proškolit.
- K plnění úkolů je třeba vytvořit vhodné pracovní prostředí.
- U větších projektů je třeba rozdělit odpovědnost mezi více pracovníků, kteří jsou odborníky na danou oblast. Tím je rovnoměrně rozložena zátěž mezi jednotlivé vedoucí pracovníky, zároveň dochází ke snížení rizika celkového selhání. Rozdělení zodpovědnosti může v projektu integrace informačních zdrojů vypadat např. takto:
  - řídicí management
  - řízení dílčích služeb/modulů
  - řízení kvality obsahu (správa informačních zdrojů)
  - řízení operačních procesů (správa serveru a aplikací, správa sítě, bezpečnost a autentikace, správa uživatelů a jejich účtů, analýza obsahu, katalogizace, podpora atd.)

### 6.1.5 Finanční zajištění projektu

#### Popis problematiky

Budování a provozování IS pro integraci informačních zdrojů je finančně náročný proces. Nezbytný je proto kvalitní finanční plán a fungující obchodní model. Je třeba finančně zajistit nejen vývoj, pořízení a implementaci IS, ale také administraci, školení, propagaci a zejména dlouhodobý provoz a rozvoj IS. Musíme finančně zajistit celkové náklady na vlastnictví, tedy všechny náklady vyžádané v průběhu celého životního cyklu provozovaného systému.

#### Praktické rady

- Pro integrační projekt je vhodné zajistit více zdrojů financování, abychom nebyli závislí na jediném zdroji. Jinou možností je navýšit běžný rozpočet knihovny o náklady spojené s projektem, to je však možné pouze na základě dohody se zřizovatelem.
- Projekt je vhodné rozčlenit do samostatných projektových etap s vlastním finančním zajištěním. Na dílčí etapy bude snadnější získat potřebné finance, např. z grantů nebo i rozpočtů.
- Životaschopnost projektu podpoříme tím, že maximálně snížíme fixní náklady, které vznikají zejm. provozní částí projektu. Úspor fixních nákladů dosáhneme např.:
  - tvorbou konsorcií (kdy na rozpočet projektu přispívá více institucí, navíc lze získat na nákup i údržbové poplatky konsorcionální slevy)
  - bezplatným využíváním software a hardware formou sponzorství (lze využít tehdy, když je projekt úspěšný a veřejně známý)
  - využíváním open source software, který je k využití zdarma, ale zároveň je k němu zajištěna podpora vývojářské a uživatelské komunity
  - využíváním hardware, software a práce formou objednané služby (ASP)
  - využitím darů (lze podpořit propagační kampaní)
  - najímáním pracovníků k realizaci konkrétním úkolů na dohodu o provedené práci nebo formou pracovního úvazku na dobu určitou
- Potřebujeme-li snížit výši počátečních investic, je možné rozložit investiční náklady do delšího období formou splátek (leasing, pronájem) nebo na základě dohody s dodavatelem. Využít lze i bankovní úvěry, pokud jsou výhodné.
- Výhodné může být investice spojovat s jinými projekty, čímž docílíme rozložení finanční zátěže mezi jednotlivé projekty (např. nákup serveru, který bude sloužit dvěma různým projektům nebo účelům).
- Významným zdrojem příjmů mohou být nejen rozpočty organizace, ale i různé typy finanční podpory vlády, ministerstev, jednotlivých krajů a měst, nadací, soukromých fondů a samozřejmě EU. Pokud není v projektu přítomen odborník na grantovou politiku, je možné využít služeb projektových a poradenských agentur, které nám s přípravou žádosti o dotace pomohou.
- Na grantové dotace se nelze spoléhat, podání žádosti neznamená automatické přiznání dotace. U některých grantových programů je navíc třeba dávat pozor, že se dotace vyplácí až zpětně, po úspěšné realizaci projektu.

## Poznámky

Obchodní model projektu je možné doplnit i o aktivní způsoby financování, přestože se jedná o nekomerční projekt. Do projektu můžeme např. zapojit vybrané komerční služby jiných firem za provizi (např. linkování z bibliografických záznamů do elektronických obchodů s knihami, CD apod.), čímž můžeme snižovat celkové náklady. Mezi další možnosti generování příjmů v rámci projektu můžeme zahrnout: prodej reklamního prostoru IS (např. bannery), zprostředkování prodeje informačních objektů (např. článků z komerčních databází placených na vyžádání), zpoplatněné informační služby (placené rešerše) apod.

### 6.1.6 Zabezpečení a správa informačního systému

#### Popis problematiky

Má-li být IS považován svými uživateli za spolehlivý a důvěryhodný, musí být zajištěna jeho bezpečnost a nepřetržitý provoz. Při projektování IS by měla být vypracována bezpečnostní politika projektu (často jako součást bezpečnostní politiky celé organizace provozovatele IS). Ty by se měla odrazit jak ve vlastním návrhu architektury IS, tak v jeho následné administraci. Především je třeba definovat, jaká data v IS chránit, a analyzovat rizika, která těmto datům hrozí. Je kriticky důležité znát co nejvíce možných typů ohrožení a ty v IS efektivně ošetřit. Jde zejm. o ohrožení útoky uživatelů, ohrožení z nedbalosti uživatelů, ohrožení haváriemi a ohrožení vinou chyb IS.

#### Praktické rady

- V každém IS by měly být dodržovány následující zásady:
  - pravidelné aktualizace a zavádění opravných balíčků software
  - nasazení antivirového software a jeho pravidelné aktualizace
  - nasazení síťového firewallu na serveru a jeho přísné nastavení
  - účinná politika hesel
  - zamezení nešifrované autentikace
  - zamezení neoprávněného rozesílání e-mailových zpráv přes IS
  - zamezení neoprávněného využívání proxy služeb
  - fyzické zabezpečení IS, zvláště serveru
  - zamezení provozu nepotřebných služeb v IS
- Systémový administrátor by měl pravidelně provádět testy na zranitelnost a napadnutelnost IS. K dispozici je řada účinných komerčních i volných nástrojů (označované např. jako file integrity checkers, intruder detection systems atd.).
- Řadu možných bezpečnostních rizik lze také odhalit nepřetržitým monitoringem pomocí speciálních nástrojů (např. na sledování modifikace práv) a pravidelnou analýzou systémových a přístupových logů. Takové nástroje by měly umět v případě ohrožení okamžitě varovat administrátora formou SMS nebo e-mailové zprávy.
- Při bezpečnostní analýze IS lze také efektivně využít nástroje útočníků, kterými můžeme ověřit odolnost proti útokům (např. password crackers, sniffers, port scanners).
- Jakákoli změna v konfiguraci hardware i software by měla být zdokumentována s přesným časovým označením.
- IS musí být účinně zálohován. Je třeba určit, co přesně se bude zálohovat, jak často, jakým způsobem a jak dlouho bude záloha uchovávána. Zálohy by měly být fyzicky odděleny od zálohovaného IS, např. pro případ živelné pohromy či krádeže zařízení. Součástí systému zálohování musí být i přesný popis procesu obnovy IS ze zálohy (viz havarijní plán popsany níže).

### 6.1.7 Havarijní plán

#### Popis problematiky

Každý IS by měl být schopen čelit všem nebezpečím, která byla definována na základě analýzy rizik. Pro každé riziko by mělo být připraveno odpovídající preventivní opatření. I přes veškerou snahu o bezpečný a plynulý chod IS však může dojít k výpadku celého IS nebo jeho části, který označujeme jako havarijní stav. Pro takový případ je nutné mít připraveny rychlé a efektivní postupy obnovy jednotlivých funkcí IS.

## Praktické rady

- Při sestavování havarijního plánu je třeba popsat pokud možno všechny možné havarijní stavy daného IS. K tomu je vhodné využít zkušenosti jiných provozovatelů, dodavatelů dílčích komponent apod. Ke každému havarijnímu stavu je třeba sestavit postup konkrétních kroků, které vedou k obnovení optimálního provozu IS.
- Za každý postup obnovy musí odpovídat konkrétní osoba. Musí být také definován způsob rychlé komunikace s dalšími pracovníky či subjekty v procesu obnovy (např. rychlé objednání a dodání náhradního dílu na základě servisní smlouvy na hardware).
- Celý proces od odhalení problému, přes identifikaci příčiny až po obnovení předchozího stavu musí být bežešvý, efektivní a co možná nejrychlejší. Toho lze dosáhnout snížením počtů operativních kroků havarijního plánu na minimum a vyřadit zbytečné mezičlánky, např. udělením rozhodovací pravomoci pracovníkovi IT pro případ havárie.
- Havarijní plán by měl být v organizaci vyhlášen jako platná směrnice, aby byl pro všechny pracovníky závazný. Příslušní pracovníci musí být proškoleni v jednotlivých postupech.
- Má-li být havarijní plán účinný, musí vždy obsahovat minimálně následující tři kroky v tomto pořadí:
  - odstranění příčiny havárie
  - obnovení nebo vyřazení poškozených komponent (postup dodání náhradních dílů, odpojení nefunkčního informačního zdroje apod.)
  - obnovení optimální funkce IS (postup reinstalace, obnova konfigurace, obnova dat)
  - přijmutí preventivních protiopatření, aby k havarijnímu stavu příště nedošlo (změna konfigurace, změna bezpečnostní politiky, důkladnější proškolení, sankce za nedbalost, zavedení QMS, změna odpovědnosti apod.)

### 6.1.8 Zavedení integračního systému mezi uživatele

#### Popis problematiky

Předpokladem úspěšného zavedení IS mezi uživatele je užitečnost, kvalita, důvěryhodnost a účinná propagace. Konkurence mezi službami postupnými na webu je dnes veliká, uživatelé si vybírají, jaké služby budou používat. Spuštění IS do ostrého provozu neznamena, že ho začnou uživatelé automaticky používat.

#### Praktické rady

- Užitečnost - integrační systém musí být pro uživatele užitečný a při práci pohodlný. Musí nabízet služby a informace s přidanou hodnotou snadno a rychle. Přidaná hodnota vzniká jako nová kvalita propojováním či obohacováním služeb, vnesením řádu do velkých objemů informací, zjednodušení rešeršních postupů, integrací kvalitních služeb vlastních i třetích stran apod.
- Kvalita – celkovou kvalitu integračního systému ovlivňují zejména: doba odezvy systému, kvalita navigace v systému (vyhledávání, menu, kategorie), kvalita obsahu, přesnost a jednotnost vyhledávání, forma reprezentace obsahu, kvalita popisu zdrojů, provozní hodiny/výpadky, aktuálnost odkazů, grafický design, intuitivnost ovládání a dodržování standardů.
- Důvěryhodnost – v IS by měly být k dispozici informace o provozovateli, kontakty s možností zasílat připomínky, náměty a dotazy. Všechny možnosti zpětné vazby a interakce uživatelé vítají. Pokud uživatelé poskytují osobní informace, měli by vždy vědět, z jakého důvodu. Garance ochrany osobních údajů a ochrany evidence aktivity uživatele by měla být samozřejmostí.
- Propagace – viz následující kapitola 6.1.9.

### 6.1.9 Propagace služeb integračního systému

#### Popis problematiky

IS pro integraci informačních zdrojů je třeba propagovat mezi stávajícími i potencionálními uživateli. Je třeba mít stále na paměti, že IS vznikl pro uživatele a uspokojování jejich informačních potřeb. Pokud IS nebudou uživatelé využívat, postrádá i sebelepší IS smysl. Cílem

propagace je získávat a průběžně udržovat uživatele. Nízká návštěvnost a nezájem uživatelů o IS je známkou selhání projektu.

### Praktické rady

- S propagací integračního systému je vhodné začít již ve fázi projektování. Uživatelé by měli vědět, že něco připravujeme, čímž se IS dostane do povědomí uživatelů ještě před jeho vlastním spuštěním.
- Ke zvyšování povědomí uživatelů o poskytovaných službách IS lze využívat metody známé z podpory prodeje: dočasné poskytování placených služeb zdarma, předváděcí akce, reference stávajících uživatelů, psaní článků do novin i odborných časopisů, vydávání propagačních letáků a brožur, přednášení na konferencích nebo v rámci výuky studentů, účast v odborných diskusích, udržování kontaktů se stávajícími uživateli (např. rozesíláním elektronických zpravodajů), publikování novinek IS na jiných serverech (např. pomocí RSS).
- Důležitou součástí propagace je reklama, která musí být originální se stručných a nejlépe vtipným sloganem. Je třeba promyslet racionální i emocionální složku reklamy. K šíření reklamy je třeba využít všechna dostupná média a reklamu opakovat.
- Je třeba organizovat pravidelná školení, zvláště když se objevují nové verze produktů, workshopy, připravovat online nápovědy a e-learningové kurzy.
- Služby IS je vhodné propagovat v širším kontextu, např. jako součást služeb univerzity nebo města.

### Poznámky

Při propagaci služeb IS pro integraci informačních zdrojů lze s úspěchem zdůraznit výhody, které přináší knihovny uživatelům obecně:

- Přístup k dokumentům, které již nejsou v prodeji (funkce archivu)
- Přístup k dokumentům, které uživatelé potřebují pouze krátkodobě, např. ke studiu (funkce půjčovny)
- Přístup k licencím placených informačních zdrojů, které by si uživatelé sami finančně nemohli dovolit (funkce klubového sdílení)
- Referenční a rešeršní služby či transformace informací (přidaná hodnota služeb)
- Elektronické dodávání dokumentů

## 6.2 ARCHITEKTURA IS PRO INTEGRACI INFORMAČNÍCH ZDROJŮ

### 6.2.1 Integrace klasického fondu

#### Popis problematiky

Klasický fond představuje velkou množinu dokumentů v analogové, především papírové podobě. Existují dva způsoby, jak klasický fond integrovat do integračního systému: buď průběžnou digitalizací do elektronické podoby nebo prostřednictvím katalogu (OPAC) a navazujících služeb elektronického dodávání dokumentů (EDD).

#### Praktické rady

- Bibliografický popis či tvorba metadat je tradiční činnost knihoven, která dobře funguje. Efektivitu je ale vhodné zvýšit sdílenou katalogizací či přebíráním záznamů.
- Uživatel se nespokojí s přístupem do katalogu či bibliografické databáze, chce získat primární dokument, a to hned a bez osobní návštěvy knihovny. Do IS je třeba maximálně integrovat službu EDD a další způsoby získávání dokumentů.
- EDD musí být co nejrychlejší a bez prodlev, proto je nutné vyřešit mikroplatby.
- OPAC knihovny se musí stát nedílnou součástí integračního systému, uživatel musí mít ke všem informačním zdrojům unifikovaný přístup bez ohledu na to, jsou-li lokální či vzdálené.
- Knihovna by měla umožnit zapojení svého OPAC i do jiných integračních projektů (např. do souborných katalogů, do oborových bran apod.). Zvýší se tak zájem o fond knihovny i mimo vlastní uživatelskou základnu.

### 6.2.2 Integrace elektronických databází a archivů

#### Popis problematiky

Elektronické databáze (obsahující primární nebo sekundární informace) a e-archivy jsou většinou vybaveny vlastním systémem pro správu báze dat a vyhledávacím strojem. Bývají opatřeny komunikačním rozhraním, které umožňuje buď vzdálené vyhledávání (typicky HTTP/XML a Z39.50) nebo naopak sklizení metadat. Odlišné vyhledávací stroje a různé způsoby indexování však způsobují, že jednotlivé databáze a e-archivy mají různé vyhledávací možnosti.

#### Praktické rady

- Bibliografické databáze mají podobnou strukturu jako katalogy klasického fondu, snadno je proto můžeme společně integrovat.
- Plnotextové databáze a elektronické archivy dokumentů jsou většinou vybaveny popisnými údaji, které můžeme integrovat na úrovni sekundárních informací s jinými bibliografickými databázemi či katalogy. Plnotextového vyhledávání by mělo být odděleno, aby uživatel vždy věděl, na jaké úrovni hledá.
- Integrace na úrovni sekundárních informací můžeme řešit buď budováním fyzické souborné báze (pomocí replikací dílčích bází či sklizení metadat) nebo virtuálně pomocí distribuovaného vyhledávání.
- Stejně jako by měl být OPAC integrován mezi ostatní elektronické databáze, měl by být i obsah nakoupených elektronických databází a e-archivů zahrnut v OPACu (v katalogu by měly být zahrnuty elektronické časopisy, e-knihy a další dokumenty obsažené v nakoupených databázích).
- Je třeba zajistit vzdálený přístup oprávněných uživatelů do databází, které jsou zabezpečené pomocí hesla či IP adresy. Integrační systém může fungovat jako proxy server.
- Uživatel by měl mít možnost pracovat s celou množinou elektronických dokumentů najednou, bez ohledu na to, v jakých databázích či e-archivech jsou obsaženy.

### 6.2.3 Integrace volného prostoru internetu

#### Popis problematiky

Internet obsahuje mnoho zcela odlišných druhů informačních zdrojů, k jejichž využití je nutné použít různé nástroje a metody. Knihovny by se měly v integračních projektech soustředit na služby, kterými mohou ve zpracování volného prostoru internetu překonat kvalitu internetových vyhledávačů typu Google. Služby, které naopak internetové vyhledávače provádějí lépe, by měl integrační systém přejímat a využívat.

#### Praktické rady

- Internetové vyhledávače mohou být do integračního systému připojeny jako jedny z informačních zdrojů a mohou doplňovat výsledky paralelního vyhledávání. Je však vhodné jim nastavit menší míru hodnocení relevance, aby se jejich výsledky řadily až za výsledky z kvalitních plnotextových zdrojů.
- Součástí IS mohou být seznamy či databáze kvalitních internetových odkazů. Ty může knihovna vytvářet sama, přejímat je z jiných projektů nebo je nakupovat jako jeden z informačních zdrojů.
- Náměty na zajímavé internetové zdroje můžeme získávat na základě kvalitních citací v odborných časopisech (na rozdíl od PageRanking Googlu), od uživatelů, z odborných diskusí či blogů, z RSS kanálů, z manuálně prováděných rešerší v rámci dané problematiky apod. Získáme tak i URL adresy zdrojů, které ještě nebyly v dalších dokumentech citovány.
- Zahrnutí nového internetového odkazu do seznamu by se mělo řídit podobnými hledisky jako u akvizice klasických dokumentů. Rozhodující je zejm. kvalita obsažených informací, jazykové hledisko, aktuálnost a úplnost informací a samozřejmě funkčnost URL.
- U internetových zdrojů je třeba rozlišovat jejich typ, protože každý vyžaduje jiné hodnocení a jiné zpracování. Rozlišujeme zejm. primární informační objekty (dokumenty) a jejich soubory (elektronické časopisy), internetové databáze a webové portály.

- Opatříme-li internetové zdroje v seznamu kvalitním popisem, umožníme uživatelům prohledávat tyto zdroje podle více hledisek, rychleji a přesněji, než to umožňují internetové vyhledávače. Díky předmětové klasifikaci bude navíc vyhledávání nezávislé na jazyku internetového zdroje.
- Kvůli vysoké proměnlivosti internetového obsahu je třeba nasadit účinné kontrolní mechanismy (průběžná kontrola platnosti URL, kontrola změny obsahu dokumentů atd.).
- Je vhodné integrovat některý z webových konzervačních archivů, aby měli uživatelé přístup i k dokumentům, které již z živého internetu nenávratně zmizely.

#### 6.2.4 Uživatelské rozhraní

##### Popis problematiky

Uživatel provádí všechny operace prostřednictvím uživatelského rozhraní. Uživatelské rozhraní je výlohou celého integračního systému, je to tlumočnick mezi uživatelem a systémem. Uživatelské rozhraní má proto poněkud schizofrenní úlohu – musí být čitelné a příjemné pro člověka, zároveň musí tvořit vstupy i výstupy stroje. U síťových aplikací, mezi které IS pro integraci informačních zdrojů patří, se stal pro uživatelské rozhraní de facto standardem webový prohlížeč. Ten svými vlastnostmi vytváří rámeček konkrétní podoby uživatelského rozhraní dané aplikace.

##### Praktické rady

- Uživatelské rozhraní by mělo být navrženo tak, aby uživatel mohl pracovat bez většího uvědomování si vlastního řešení rozhraní. Uživatel by se měl soustředit na svoji práci, nikoli na rozhraní.
- Uživatel by měl být schopen předvídat chování jednotlivých prvků rozhraní, a to jednak podle vizualizace prvků a jednak na základě zkušeností získaných v jiných aplikacích. Všechny prvky by měly být samovysvětlující.
- Na všech místech v IS by měly být termíny a označení používány konzistentně. Jednotnost pojmenování příkazů, ovládacích prvků a komponent by měla být zachována i v tištěných příručkách, nápovědách a dalších propagačních materiálech.
- V uživatelském rozhraní by neměly být použity odborné výrazy, kterým laický uživatel nemusí rozumět. Příkazy a odkazy by měly vyjadřovat stav, kterého bude aplikací dosaženo. Např. „Získejte v jiné knihovně“ místo nevhodného výrazu „Meziknihovní služba“ nebo „Nalézt článek“ místo „Hledat v EBSCO“.
- Grafický design má podporovat a zpřehledňovat rozložení prvků na obrazovce, nesmí být pro práci uživatele přítěží.
- Uživatelské rozhraní musí být jednoduché a srozumitelné. Musí uživatele vést.
- Uživatelské rozhraní by mělo respektovat možné scénáře pracovních postupů. Uživatel by se neměl nikdy ocitnout ve slepé uličce, vždy by měl mít možnost pokračovat v započatém procesu. Hlášky typu „Nic nebylo nalezeno.“ bez možnosti dalších kroků jsou nepřijatelné.
- Uživatelské rozhraní by nemělo nutit uživatele provádět operace v určitém pořadí, pokud to není nezbytně nutné.
- Uživatel musí po každé vykonané akci v rozhraní získat odpovídající zpětnou vazbu.
- Uživatel by měl mít možnost upravovat a udržovat svůj pracovní prostor podle svých požadavků.
- Veškeré výjimky by měly být ošetřeny pomocí srozumitelných chybových hlášek.
- Uživatelské rozhraní musí být provedeno v souladu s bezpečnostními požadavky IS.
- Uživatelské rozhraní musí odpovídat platným standardům použitých technologií, nemělo by záviset na platformě klientského zařízení.
- Řadu slabin a nesrozumitelných vazeb rozhraní můžeme odhalit přes ostrým spuštěním IS díky spolupráci betatesterů z řad uživatelů. Pozorováním jejich chování při práci se systémem a shromažďováním jejich připomínek získáme cenné údaje pro analýzu testované verze IS.

##### Poznámky

V ČR je přibližně 100 tisíc osob s těžkým zrakovým postižením<sup>30</sup>. IS by měl být technologicky uzpůsoben také pro ně. Doporučení pro vývoj aplikací pro zrakově postižené uživatele definuje

např. pravidla W3C WAI (Web Accessibility Initiative) a WCAG 2.0 (Web Content Accessibility Guidelines).

### 6.2.5 Orientace na uživatele a spolupráce s dalšími službami

#### Popis problematiky

Integrační systém umožňuje knihovnám překonávat svoje vlastní bariéry a umožňuje jim zapojit se do skutečně kooperativního prostředí internetu, ve kterém fungují tisíce zajímavých služeb. Knihovní služby se tam mohou stát plnohodnotnou součástí internetových služeb pro získávání informací. Proto je však třeba uzpůsobit prostředí integračního systému knihovny potřebám a zvykům uživatelů internetu. Jde především o možnost přizpůsobovat si své pracovní prostředí, možnost dále zpracovávat a třídit získané informace, podílet se na tvorbě obsahu, vyjadřovat svoje názory a sdílet informace s ostatními uživateli. Velký důraz je kladen především na samoobslužnost služeb a dynamičnost obsahu, díky níž prochází IS nepřetržitou změnou.

#### Praktické rady

- Uživatelské prostředí by mělo umožňovat dynamickou a samoobslužnou konfiguraci individuálně pro každého uživatele. Jednotlivé komponenty rozhraní by měl mít uživatel možnost zapnout nebo vypnout, nastavit si jejich chování a vzhled.
- Pokud to zaměření IS umožňuje, je vhodné pro uživatele vytvářet jejich vlastní uživatelské schránky, které slouží jako úložiště vybraných záznamů, dokumentů, předdefinovaných rešeršních dotazů, odkazů či výstřižků elektronických textů. Uživatelé také vítají možnost, kdy mohou vybrané složky svých uživatelských schránek sdílet s ostatními uživateli. To má velké uplatnění v akademické i zájmové sféře uživatelů.
- Kromě informačních zdrojů, které knihovníci považují za cenné a kvalitní, je vhodné do celkové nabídky zdrojů zapojit i malé, okrajové zdroje, které podle průzkumů uživatelé používají. Uživatelé pak nemusí tyto zdroje využívat mimo náš IS a na základě výběru si mohou obohatit rešerše i o okrajové, ale pro ně zajímavé zdroje.
- Uživatelé často rádi vytvářejí uživatelské skupiny, zejm. podle studijních, zájmových či pracovních struktur. IS by měl tvorbu takových skupin podporovat, zejm. z hlediska sdílení a uzpůsobené nabídky.
- Operace, u kterých se očekává, že je uživatelé budou provádět opakovaně, je vhodné rozšířit o možnost automatického spouštění. Jsou to především pravidelně spouštěné dotazy SDI, rozesílání informační avíz a upozornění formou e-mailu, RSS kanály, automatické údržby obsahu uživatelských schránek apod.
- Zejm. u oborově zaměřených IS je vhodné zvážit možnost účasti uživatelů na tvorbě obsahu IS. Možných forem je mnoho. Od možnosti komentovat a obohacovat již existující informační objekty (články, záznamy, fotografie), přes tvorbu diskusních příspěvků a blogů až po rychlé vytváření obsahu webu s možností zpětné editace (přetváření obsahu na principu wiki).

### 6.2.6 Využití všech komunikačních kanálů

#### Popis problematiky

Přestože jsou dnes IS orientovány hlavně na web, je přínosné podporovat i další komunikační kanály, pomocí nichž mohou uživatelé IS využívat. Jsou tak odbourány mnohé technologické bariéry a IS může využít opravdu každý a za každé situace.

#### Praktické rady

- Možnými způsoby komunikace mezi IS a uživateli jsou:
  - SMS komunikace (např. pro výpis výpůjček a prodloužení, zasílání avíz atd.)
  - WAP nebo webový přístup uzpůsobený pro malá mobilní zařízení
  - elektronická pošta
  - fax (alternativa elektronické pošty)
  - telefonické spojení (pomocí automatu nebo pracovníka služby)
  - poštovní doručování (zasílání tištěných výstupů)

## Poznámky

Každý podporovaný komunikační kanál by měl v maximální míře využívat automatizované procesy, i když se jedná o kanál neelektronický. Např. při poštovním doručování může IS tisknout adresní štítky z databáze uživatelů. S těmito funkcemi je třeba počítat již při návrhu IS.

### 6.2.7 Volba produktů a technologií

#### Popis problematiky

Výběr software i hardware by měl odrážet cíle a jednotlivé požadavky projektu. Ať již vyvíjíme proprietární software, pořizujeme hotová řešení nebo tyto způsoby kombinujeme, vždy je třeba určit, jaké technologie budou použity, zejména s ohledem na připojované externí informační zdroje a informační okolí. V některých případech může být výhodnější konkrétní funkcionalitu objednat jako službu než ji vyvíjet vlastními silami a zajišťovat její provoz.

#### Praktické rady

- Vždy bychom měli pečlivě zvážit, jestli není možné konkrétní funkcionalitu IS zajistit externě formou služby, např. pronájemem zařízení nebo externím provozem. Výhodné jsou zejm. služby ASP (application service providing), tedy provoz aplikace na dálku přes internet. ASP často vychází levněji než součet investic do vlastního hardware, software, vývoje a provozních nákladů (celkové náklady na vlastnictví). Pokud by však ASP znamenalo ztrátu klíčových výhod vlastního provozu, měli bychom se rozhodnout pro zajištění funkcionality IS vlastními prostředky.
- Požadavky na software zřídka kdy pokryje jeden produkt nebo konkrétní řešení od jednoho výrobce. Vhodnější je hledat modulární řešení různých výrobců s důrazem na vzájemnou kompatibilitu.
- Vlastní proprietární řešení je vhodné nasazovat až v případě, kdy požadavky nespĺňuje žádný z dostupných produktů na trhu.
- Důvodem k proprietárnímu řešení může být i nedostatek financí na nákup komerčního software. Stále častěji jsou ale k dispozici kvalitní open source aplikace, které nejenže šetří investice, ale zároveň umožňují modifikace kódu aplikace.
- Jednotlivé moduly řešení musí být nahraditelné a flexibilní na změny ostatních modulů. Nezbytností je proto podpora průmyslových a oborových standardů, zejm. u rozhraní, které zajišťují komunikaci IS s okolím (výměnné formáty, exportní funkce apod.).
- Vždy je vhodné porovnávat různé alternativy řešení. U každé alternativy je třeba hodnotit celkový dopad na IS a na celkové náklady na vlastnictví z dlouhodobého hlediska. Hledání nejlevnějších alternativ bez ohledu na výkon může vést k nefunkčnímu IS.
- V jednotlivých modulech IS, které dodavatelé nabízejí, může být obsaženo velké množství technologií, se kterými nemusí mít projektový tým přímé zkušenosti. Při posuzování nabídek je proto vhodné využít služeb poradenských firem.

## 6.3 AKVIZICE INFORMAČNÍCH ZDROJŮ DO INTEGRAČNÍHO SYSTÉMU

### 6.3.1 Důraz na primární dokumenty

#### Popis problematiky

Převážná většina uživatelů potřebuje pro uspokojení svých informačních potřeb získat primární nebo faktografické informace. Bibliografické informace vystupují pouze jako informační pomůcky při vyhledávání a získávání vlastních informačních objektů. Knihovníci často inklinují k přeceňování významu bibliografických informací a zapomínají na to, že informační proces musí být korunován dodáním primárního dokumentu či referenční informace, jinak celý proces postrádá smysl.

#### Praktické rady

- Do nabídky pro koncové uživatele bychom neměli v integračním systému začleňovat čistě bibliografické nebo popisné databáze, pokud nemají přímo v IS vazby na primární objekty



nebo služby EDD či MVS. Bibliografické informace bez dalších vazeb jsou pro koncové uživatele bezcenné a v nabídce matoucí.

- Katalogy knihoven je vhodné integrovat pouze v případě, že může uživatel popisované dokumenty získat formou výpůjčky nebo EDD, a to s možností objednání přímo v integračním systému.

### 6.3.2 *Důraz na elektronickou podobu dokumentů*

#### **Popis problematiky**

Integrační systém by měl v nabídce upřednostňovat ty zdroje, které nabízejí primární informační objekty v elektronické podobě. Vždy bychom měli pamatovat na ideál dostupnosti „tady a teď“ (požadované dokumenty na obrazovce kdykoli a odkudkoli), kterého lze dosáhnout pouze u elektronických forem informačních objektů.

#### **Praktické rady**

- U velmi žádaných dokumentů v klasické podobě bychom měli uvažovat o postupné digitalizaci, aby mohly být dostupné elektronicky. Je třeba pečlivě zvažovat, co digitalizovat, protože je to jeden z nejnákladnějších procesů zpracování dokumentů.
- Soustředit bychom se měli na akvizici původních elektronických prací. Mnoho publikací je vydáváno simultánně klasicky i elektronicky. Pro ukládání souborů elektronických dokumentů je však třeba zvolit vhodné úložiště, např. digitální repozitář, který následně zapojíme do integračního systému.
- Významným zdrojem elektronických dokumentů je internet. Nestálost a dynamičnost změn dokumentů na internetu mohou řešit různé konzervační projekty, které i přes velké právní problémy na národních i lokálních úrovních selektivně archivují internetový obsah.
- Při akvizici elektronických dokumentů bychom měli posuzovat jejich technický formát s ohledem na možnost budoucí transformace, protože dokumentové technologie se rychle mění a za několik let nemusí existovat způsob, jak elektronické dokumenty správně zobrazit a interpretovat.
- Před akvizicí elektronického zdroje bychom měli vždy posoudit možnosti jeho prohledávání, zejm. vlastnosti vyhledávacího stroje a kvalitu selekčních údajů obsažených informačních objektů.

### 6.3.3 *Pokrytí*

#### **Popis problematiky**

Stejně jako u tradiční dokumentové akvizice v knihovnách, je třeba při tvorbě nabídky informačních zdrojů v integračním systému dbát na optimální tematické pokrytí. Podle zaměření projektu a potřeb uživatelů musíme pečlivě a adekvátně pokrývat konkrétní vědecké obory a předměty uživatelských zájmů. Pomoci nám mohou standardní akviziční metody a u agregátorů také analýzy překryvu. Celkové portfolio zdrojů by však mělo zajišťovat zastupitelnost jednotlivých zdrojů, aby v případě výpadků jednoho zdroje neutrpěla celková informační nabídka a pokrytí.

#### **Praktické rady**

- Před výběrem a nákupem článkových databází musíme provést analýzu jejich vzájemného překryvu, tj. určit, jaké tituly jsou v databázích duplicitní a jaké jsou naopak jedinečné. Optimální kombinací databází můžeme dosáhnout nejširší možné nabídky titulů za minimum peněz.
- Vhodnou metodou, jak popsat a řídit pokrytí a složení nabídky informačních zdrojů v integračním systému, je metoda konspektu. Díky široké mezinárodní podpoře konspekt navíc umožňuje srovnávání informačního portfolia s jinými projekty a IS.

### 6.3.4 *Volba optimálního typu akvizice*

#### **Popis problematiky**

S příchodem elektronických informací se začaly objevovat další typy akvizice informačních zdrojů, než jakým je tradiční nákup známý u klasických dokumentů. Jedná se především o různé formy „pronájmu“ (např. oprávnění k přístupu do databáze po dobu jednoho roku) či

selektivního nákupu dílčích položek v nabízeném celku (např. nákup konkrétního článku bez nutnosti zakoupit celý elektronický časopis nebo databázi). Vždy je třeba zvážit, jaká varianta je pro daný informační zdroj finančně i dlouhodobě nejvýhodnější. Rozhodující je předpokládané využití.

### Praktické rady

- Nákup fyzických i elektronických dokumentů má význam v případě, že je chceme mít k dispozici trvale, a to i z dlouhodobého hlediska. Nákupem se dokumenty stávají naším majetkem, je však třeba bedlivě sledovat licenční oprávnění a autorská práva, aby bylo možné je poskytovat uživatelům IS.
- Pronájem na základě licence je běžný způsob, jak zajistit přístup do elektronických databází. Principiálně se podobá paušálnímu hrazení služeb, vyplatí se proto pouze při vysokém využití, kdy se cena dílčího objektu snižuje s rostoucím počtem požadavků nebo v případě rychlého zastarávání informací.
- Selektivní nákup informačních objektů až v okamžiku, kdy ho uživatel požaduje, je vhodná volba akvizice u informačních zdrojů, které uživatelé využívají zřídka, ale které pro ně přesto představují velice cenné zdroje. Selektivní nákup můžeme výhodně propojit se službami EDD nebo ho realizovat na základě seznamu deziderát.
- Podklady pro rozhodování o volbě typu akvizice nám mohou poskytovat analýzy realizovaných vyhledávání v IS – přehledně vidíme, o jaká témata je největší zájem (nejfrekventovanější výrazy) a co v nabídce chybí (výrazy vracející nulové nebo nízké výsledky).
- Měli bychom sledovat využívání již zapojených informačních zdrojů, zejm. z hlediska vyhledávání a stahování informačních objektů. V případě malého zájmu uživatelů je nutné zdroj zpropagovat a pokud nízké využití nadále přetrvává, měli bychom zvážit případné vyřazení zdroje z nabídky. Sledování využití jednotlivých zdrojů by mělo být umožněno i u agregátorů, tj. při vnořování do integrujícího celku.
- Abychom mohli porovnávat využití s jinými institucemi, je vhodné sledovat statistické ukazatele podle rozšířených standardů, např. normy COUNTER, kterou dnes podporuje většina světových vydavatelů a knihoven.

### 6.3.5 Garance důvěryhodnosti informačního zdroje

#### Popis problematiky

V oblasti odborných informací je klíčovou vlastností informačního zdroje jeho důvěryhodnost. Důvěryhodnost by měla být dána mírou ověřitelnosti sdělovaných informací a odkazy na citované prameny. Uživatel však dnes nemá možnost si veškeré informace přímo ověřovat, musí se proto spoléhat na tzv. garanci důvěryhodnosti. Při akvizici informačních zdrojů bychom proto měli vybírat zdroje, jejichž producenti mohou zaručit spolehlivost a ověřitelnost obsažených informací. Tuto selekci sice může někdo považovat za cenzuru některých informačních zdrojů, uživatel má však stále možnost takové zdroje využít mimo integrační systém. Akvizice informačních zdrojů podle důvěryhodnosti jsou pro většinu uživatelů jistým přínosem.

#### Praktické rady

- U vědeckých časopisů bývá často uváděn tzv. impact faktor, tedy ukazatel vlivu časopisu na daný vědní obor měřený na základě citovanosti. Časopisy s vysokým impact faktorem publikují informace, které mají zásadní vliv na další vědecké práce a jsou proto vysoce citované. Zároveň jsou v nich uvedené poznatky a informace přísně ověřovány okruhem odborníků. Můžeme využít mnoho informačních zdrojů, které informaci o impact faktoru časopisů přinášejí, např. Journal Citation Report od Thomson ISI.
- Cenné informace o důvěryhodnosti informací přinášejí recenze. Vědecky recenzované publikace jsou zárukou vysoké kvality a důvěryhodnosti informací.
- Důvěryhodnost může být indikována i „dobrou pověstí“ vydavatelů či uznávaných autorů.
- Jednotlivé dokumenty by měly vždy obsahovat nějakou formu podpory tvrzení, např. citace, odkazy na argumentačně podobné dokumenty nebo důkazy. Při akvizici můžeme namátkově přítomnost těchto prvků prověřit.
- Prohlášení o garanci důvěryhodnosti by mělo být uvedeno v nákupní licenci či licenční smlouvě informačních zdrojů. Před nákupem bychom se měli s tímto prohlášením

seznámit. Pokud se producent zříká jakékoli odpovědnosti za obsažené informace, neměli bychom zdroj do integračního systému zapojovat.

- Při paralelním vyhledávání a souhrnném zobrazování výsledků by měl být vždy uveden originální zdroj dané informace. Uživatel tak bude mít možnost zpětně ověřit důvěryhodnost informace.
- Informační zdroje provozované anonymně či s autorsky neoznačenými dokumenty bychom měli v integračním systému eliminovat, protože nejsou zárukou důvěryhodnosti. Pokud se rozhodneme je zapojovat, měly by být v tomto smyslu označeny, aby uživatel věděl, že má být při akceptaci tvrzení opatrnější.

## 6.4 POPIS INFORMAČNÍCH ZDROJŮ

### 6.4.1 *Potřeba jednoznačného, globálního a trvalého identifikátoru zdroje*

#### Popis problematiky

Při integraci informačních zdrojů je třeba mít nad zdroji trvalou kontrolu. Ty však mohou měnit v průběhu času název, síťovou adresu, provozovatele, často dokonce kompletní obsah, čímž mohou velmi znesnadňovat integrační snahy. Je proto vhodné využívat jednoznačnou, trvalou a mezinárodně platnou identifikaci zdrojů.

#### Praktické rady

- Není vhodné zavádět vlastní identifikátory informačních zdrojů. Trend naznačuje jasný sklon k identifikátorům typu URI, které identifikují zdroj jako takový i jeho umístění (URN a URL). Příkladem je kombinace ISSN časopisu a jeho URL adresa.
- Před nasazením konkrétního URN je třeba ověřit, je-li k dispozici příslušná interpretační služba (tzv. resolver). Příkladem jsou např. agentury ISSN zajišťující přidělování a rozpoznávání kódů ISSN.
- Producenti software pro integraci zdrojů, který se rozhodneme používat, také často zavádějí proprietární identifikátory v rámci svých znalostníchází – při výměně software však nastane vážný problém s převoditelností popisů a identifikace v novém IS.
- Vývoj jednoznačných identifikátorů je velmi bouřlivý, je proto nutné neustále sledovat vývoj a trendy zejm. na úrovni národních knihoven. Ze soudobých URN identifikátorů je vhodné sledovat vývoj ISBN, ISSN, DOI, ISRN, ISMN, ISAN, ISRC a SICI, v oblasti URL pak vývoj řešení PURL.
- Jako výhodné se ukazuje označovat zdroje nejen vlastním identifikátorem zdroje ale i identifikátorem jeho provozovatele. K tomu lze využít platný standard ISIL (International Standard Identifier for Libraries and Related Organizations, ISO 15511).

### 6.4.2 *Popis zdrojů využitelný běžným uživatelem*

#### Popis problematiky

Knihovníci mají tendenci zhotovovat precizní a bohaté popisy informačních objektů, uživatel však potřebuje s IS pracovat rychle a přehledně bez nutnosti pročítání dlouhých popisných textů. Nabízet informační zdroje v jednoduchém uživatelském rozhraní při zachování precizního vyhledávání je nelehký úkol.

#### Praktické rady

- Při návrhu způsobu popisu připojovaných informačních zdrojů a jejich prezentace v IS často pomůže střízlivý nadhled a srovnání s odlišným ale analogicky podobným systémem. Na informační zdroje se můžeme např. dívat jako na zboží v e-shopu: popis má vyjadřovat, co obchod nabízí (evidence zboží), jaké klíčové vlastnosti má nabízené zboží (jmenný a věcný popis jednotlivých výrobků) a jednoznačně označovat vlastnosti, podle kterých můžeme zboží vyhledávat, porovnávat, třídit a filtrovat. Více informací je pro koncového uživatele nadbytečné a působí na něho rušivě.
- Při návrhu popisného schématu je vhodné využít rozšířené standardy pro popis informačních zdrojů v terciálních databázích, tj. DC metadata vyjádřená v RDF nebo MARC 21. To umožní mimo jiné i výměnu a přejímání popisů z jiných IS a ušetřit tak mnoho vlastní práce.

- Popis může úspěšně sloužit také k propagaci informačních zdrojů: může označovat, jaký zdroj je v kolekci nově, jaký zdroj je mezi uživateli oblíbený atd. Takovéto informace je vhodné v uživatelském rozhraní prezentovat vizuálně výrazným způsobem.
- Popis by měl zcela jednoznačně identifikovat, o jaký typ informačního zdroje se jedná z pohledu zmíněného vnořování informací (viz kapitola 2.6.2). Uživatel tak může poznat, jaký typ informací mu informační zdroj vrátí. To je zvláště důležité u paralelního vyhledávání, kdy mohou být souběžně prohledávány jak plné texty, bibliografické záznamy či odkazy na jiné portály, což může být ve sloučených výsledcích pro uživatele značně matoucí. Typ informačního zdroje by měl být jasně zobrazen jak v evidenční části IS, tak ve výsledcích vyhledávání.
- Součástí popisu mohou být dynamicky generované (a tedy vždy aktuální) údaje o využití či popularitě zdroje, jako je např. uživatelské hodnocení, kvantita využití, překryv obsahu s jinými zdroji apod.

### 6.4.3 Předmětová klasifikace zdrojů

#### Popis problematiky

Většina IS pro integraci informačních zdrojů je předmětově orientována nebo alespoň podporuje výběr informačních zdrojů podle předmětových kategorií. Toto zaměření má svůj důvod – uživatelé velmi často vyhledávají informační zdroje podle předmětu jejich zájmu.

#### Praktické rady

- Je třeba zvolit srozumitelné a jednoduché klasifikační schéma. Pro desítky integrovaných zdrojů nemá smysl zavádět klasifikační systém se stovkami kategorií.
- Nejlépe se osvědčily systémy se dvěma hierarchickými úrovněmi (kategorie a podkategorie), např. konspekt či různé oborově specifické systémy.
- Klasifikační schéma může být také řešeno s respektem k prostředí, ve kterém je IS nasazen. Např. může kopírovat rozdělení podle fakult univerzity či vědeckých ústavů instituce.
- Uživatelé by měli mít možnost zařazovat zdroje do vlastních předmětových kategorií v rámci svých uživatelských schránek (např. podle přednášek či osobních zájmů).
- Zajímavou alternativou mohou být klíčová slova vytvářená samotnými uživateli (tzv. folksonomie) vyjadřující obsah zdroje.

## 6.5 VYHLEDÁVÁNÍ V INFORMAČNÍCH ZDROJÍCH

### 6.5.1 Důraz na přímé vyhledávání primárních informací

#### Popis problematiky

Postupná přeměna informačních zdrojů do elektronické podoby přináší uživatelům možnost vyhledávat primární informace přímo, bez nutnosti používat informační pomůcky, jako jsou samostatné bibliografické databáze, seznamy časopisů apod. Moderní digitální sbírky, elektronické archivy a článkové databáze obsahují vlastní digitální objekty i jejich popisy, takže uživatel může vyhledávat podle popisných informací i plnotextově přímo ve zdroji. Úloha sekundárních informačních pomůcek postupně klesá.

#### Praktické rady

- Nenuťme uživatele používat vyhledávací pomůcky, které nepotřebují. Ideálem uživatelů je internetový vyhledávač a elektronický obchod. Uživatel očekává primární informaci na jedno kliknutí.
- Je třeba podporovat oba základní vyhledávací scénáře: objevování (uživatel začíná od tématu nebo klíčového slova a ještě neví, jaké dokumenty vlastně hledá) a hledání známého (uživatel začíná citací nebo abstraktem a chce se dostat na plný text).
- Vyhledávací pomůcky postavené pouze na sekundárních informacích mají význam pouze u informačních zdrojů, které nejsou v elektronické podobě, a které zpřístupňujeme pomocí výpůjčních služeb nebo služeb EDD.

- Časopisecké tituly dnešní uživatele zajímají většinou pouze z hlediska důvěryhodnosti informací. Uživatelé hledají převážně články, méně již časopisy. Dříve museli uživatelé nejprve získat časopis, aby se dostali k článku, ale dnes mohou článek vyhledat přímo v agregačních článkových databázích. Pro uživatele je proto často nepodstatné, v jakém časopise byl článek publikován, pokud je zajištěna jeho důvěryhodnost.
- Stejně jako u elektronických obchodů rozhoduje cena, rychlost dodání a kvalita služeb, nikoli informační zdroj jako takový. Seznamy časopisů jako výchozí rozhraní pro vyhledávání článků jsou pro většinu uživatelů matoucí a neefektivní.

### 6.5.2 Distribuce dotazů z integračního systému do dílčích IS

#### Popis problematiky

Každý informační zdroj, který integrujeme do našeho IS distribuovaně, může používat více či méně specifický dotazovací jazyk a různé typy, strukturu i obsah vyhledávacích indexů. Tyto zdroje mohou navíc využívat databázové stroje velmi odlišných vlastností s výčtem omezení.

#### Praktické rady

- Je vhodné definovat minimální sadu vyhledávacích bodů, které musí každý připojovaný zdroj podporovat. Tyto body by měly odpovídat údajům, podle kterých uživatelé nejčastěji vyhledávají. Pokud zdroj požadavku nevyhoví, neměl by být až na zvláštní případy integrován.
- Producenty je vhodné motivovat k podpoře vyhledávání podle minimální sady vyhledávacích bodů např. vyhlášením národních vyhledávacích profilů, jak je známe např. v případě protokolu Z39.50. Profily by měly definovat nejen výčet vyhledávacích bodů, ale i strukturu dat a způsob jejich indexování.
- Velkým přínosem je používání řízených autoritních hesel pro selekční údaje, pokud jsou pro dané vyhledávací body k dispozici. Informační zdroje, které používají společný systém autorit, budou mít harmonizované selekční údaje a vyhledávání bude přesné napříč všemi zdroji.
- Respektují-li producenti některý ze standardizovaných vyhledávacích profilů, je snadné v integračním systému nastavit poměrně přesné konverze dotazů a vyhledávacích termů pro každý typ informačního zdroje.
- Užitečnou pomůckou mohou být vzorové rešeršní scénáře pomocí nichž je možné připojovaný zdroj otestovat a zhodnotit jeho vyhledávací možnosti. Porovnáním vyhledaných výsledků v nativním rozhraní zdroje a výsledků vrácených z integračního systému získáme přesné informace o problémech při distribuci dotazu.

#### Poznámky

Při distribuovaném vyhledávání je třeba pravidelně sledovat statistiky a logy, protože souběžné využívání informačních zdrojů může značně vytěžovat dostupné licence a způsobovat nedostupnost zdrojů.

### 6.5.3 Uživatelsky přívětivé vyhledávání

#### Popis problematiky

Cílem vývojářů vyhledávacích strojů je maximální rychlost, přesnost a robustnost vyhledávání. To se děje složitými algoritmy na systémové úrovni IS. Koncový uživatel však musí být od této vrstvy odstíněn, aby mohl co nejpohodlněji pracovat v prostředí, které odpovídá jeho potřebám.

#### Praktické rady

- Uživatel musí v celém procesu vyhledávání vědět, v jaké fázi vyhledávání je, kam směřuje a jaké výsledky může očekávat. Celý proces by měl umožňovat snadnou cestu k primárnímu dokumentu.
- Uživatel by měl mít možnost, ale ne povinnost, vybrat při zadávání dotazu, v jakých konkrétních zdrojích vyhledávání proběhne. Pokud uživatel nechce tuto možnost využít, měly by být předem vybrány výchozí zdroje, ve kterých bude IS vyhledávat. Alternativou může být volba „prohledat vše“, která má však smysl pouze u IS s malým počtem připojených zdrojů.

- Nabídka zdrojů pro vyhledávání musí být přehledně členěna. Uživatele nejvíce odrazují dlouhé seznamy.
- Na začátku nabídek by měly být nabízeny zdroje, které jsou nejoblíbenější a které proto pravděpodobně nejlépe odpovídají informačním potřebám uživatelů. Zrovna tak můžeme vyhledaným objektům z těchto zdrojů zvýšit míru relevance, aby se zobrazily ve výsledném seznamu na předních místech.
- Abecední soupisy informačních zdrojů a časopisů již uživatelé téměř nepoužívají, není tedy třeba je vytvářet. Je to stejně zbytečné jako vytvářet abecední seznam webových stránek.
- Uživatel by měl mít vždy možnost přejít do nativního rozhraní informačního zdroje, které může nabízet pokročilejší možnosti vyhledávání.
- Podle statistických analýz<sup>5</sup> vytvářejí uživatelé více než 70% dotazů pomocí klíčových slov a většina z nich je směřována na články. Vyhledávací rozhraní a celý vyhledávací proces by měl tento fakt respektovat.
- Je třeba vyhovět příznivcům Google i náročnějším uživatelům – k dispozici by mělo být jednoduché jednopolíčkové vyhledávání i pokročilé vyhledávání s možností kombinací podmínek a filtrování.
- Mnoho uživatelů netuší, co je to integrační systém a paralelní vyhledávání. Projekt je třeba propagovat zdůrazněním výhod ve vyhledávání. Je však třeba i otevřeně upozornit na nevýhody, mezi které patří např. nižší rychlost distribuovaného vyhledávání, než kterou přináší IS s lokálními indexy, např. Google.

#### 6.5.4 Paralelní vyhledávání

##### Popis problematiky

Paralelní vyhledávání představuje složitý proces, který vyžaduje precizní konfiguraci. Je třeba sledovat všechny faktory, které kvalitu a rychlost paralelního vyhledávání ovlivňují.

##### Praktické rady

- Při konfiguraci každého informačního zdroje pro paralelní vyhledávání je nezbytné provést následující kroky:
  - u posuzovaného zdroje poznat vlastnosti vyhledávacího stroje a strukturu indexů
  - vyhodnotit zdroj jako vhodný k paralelnímu vyhledávání
  - nakonfigurovat mapování a konverze dotazů
  - vyřešit případné jazykové rozdíly, zejm. u klíčových slov a předmětových hesel
  - nastavit zpracování a případné konverze vrácených výsledků (deduplikace apod.)
  - nastavit proces slučování s výsledky z jiných zdrojů, případně další zpracování
  - ošetřit případné časové prodlevy při vyhledávání
  - ošetřit případné přetížení a výpadky zdroje
- Při paralelním vyhledávání by se měly na výstupu vracet výsledky na stejné analytické úrovni (např. neslučovat do jedné skupiny články i časopisy) nebo by měly být rozdíly úrovně či typu dokumentu výrazně označeny.
- Jednou z možností, jak analytickou úroveň sjednotit, je budování lokálních indexů primárních dokumentů bez ohledu na stupeň vnoření v dílčích zdrojích. Vyhledávání pak může probíhat nad těmito indexy na úrovni primárních dokumentů. Jako druhou možnost můžeme zdroje rozdělit podle jednotlivých úrovní vnoření a vyhledávat ve vzniklých skupinách odděleně.

## 6.6 PREZENTACE DAT V INTEGRAČNÍM SYSTÉMU

### 6.6.1 Metadata vyhledaných informačních objektů

#### Popis problematiky

Jednotlivé informační zdroje, které v IS integrujeme, využívají k popisu obsažených objektů často různé popisné formáty nebo používají popisné formáty odlišně. Popisné formáty se navíc mohou v průběhu životního cyklu zdroje měnit (např. přechod z UNIMARC na MARC 21).

Integrační systém musí být schopen heterogenitu popisu informačních objektů řešit, aby uživatel na výstupu mohl pracovat s jednotnou množinou výsledných popisných záznamů.

### Praktické rady

- Nejběžnějším způsobem překonávání heterogenity popisných formátů je jejich konverze do vnitřního formátu integračního systému. Dávkové offline konverze (typické pro centralizovaná integrační řešení jako jsou souborné katalogy) umožňují díky větší časové toleranci složitější konverzní algoritmy a poskytují také prostor k harmonizaci a obohacování záznamů o další údaje z dalších informačních zdrojů. Ke klasickým bibliografickým údajům tak může být např. připojen obrázek obálky knihy, recenze, anotace atd., přestože tyto doplňující údaje nebyly v původním zdroji k dispozici. Online konverze jsou typické pro paralelní vyhledávání; jsou navíc z pohledu celkových nákladů levnější, protože jsou konvertovány pouze záznamy, se kterými chce uživatel právě pracovat.
- Při výběru vnitřního popisného formátu IS, do kterého budou konvertovány všechny objekty v jiných formátech, je vhodné vybrat ten, ve kterém jsou objekty připojovaných informačních zdrojů popisovány nejčastěji. Díky tomu nebudeme muset v mnoha případech popisné údaje vůbec konvertovat.
- Při práci s mnoha popisnými formáty je kriticky důležitá potřeba jednoznačných identifikátorů informačních objektů. Stejný objekt může být v různých informačních databázích popsán v různých formátech nebo odlišnými způsoby, na výstupu je však třeba identifikovat, že se jedná o tentýž objekt. Jednoznačná identifikace je předpokladem pro deduplikaci objektů i pro tvorbu jednoznačných vazeb a linkování. Při akvizici informačních zdrojů dáváme proto přednost těm, ve kterých jsou objekty opatřeny standardizovanými jednoznačnými identifikátory.

### 6.6.2 Zobrazení výsledků a možnosti dalšího zpracování

#### Popis problematiky

Celý proces vyhledávání má smysl pouze tehdy, když uživatel dostane na obrazovku smysluplné výsledky prezentované v přehledné a srozumitelné formě. Uživatel musí být schopen vrácené výsledky zpracovat a vědět, jak může výsledky dále využít.

#### Praktické rady

- Není třeba se držet pouze zaběhlých forem prezentace výsledků, jako jsou citační záznamy či tabulkové zobrazení výsledků. Uživatelé stále více tíhnou k vizualizaci. IS může generovat sémantické a interaktivní mapy výsledků, dynamické seznamy obsažených témat (označované jako topics či clusters) nebo fasety. Vše by však mělo být volitelné. Jako inspirace mohou sloužit experimentální vyhledávače kartoo.com, grokker.com, clustry.com či mooter.com.
- Vrácené výsledky by neměly být konečnou fází vyhledávání. Systém by měl vždy nabízet možnosti, jak pokračovat v práci – ať již zpřesňováním dotazu, dalším zpracováním výsledků nebo přechodem na navazující služby.
- Velice užitečným nástrojem může být seznam relevantních odkazů pro další vyhledávání. Tím mohou být obsahy polí již vyhledaných záznamů (např. hledej další díla tohoto autora), návrhy jiných gramatických tvarů pro případ překlepů (měli jste na mysli ...?) nebo odkazy na objekty, které v minulosti hledali uživatelé provádějící stejné vyhledávání (uživatelé zajímající se o ... také hledali ...).
- K lepší orientaci ve výsledných množinách vyhledaných objektů mohou sloužit nejrůznější pomůcky, od vyznačené míry relevance, přes uživatelský rating až po recenze a komentáře uživatelů.
- Základním požadavkem na prezentaci výsledků je jednoduchost a intuitivnost, což souvisí s celkovým pojetím uživatelského rozhraní. I když nabízíme různé alternativní způsoby prezentace výsledků, musíme vždy nabízet i klasické způsoby, na které je většina uživatelů zvyklá.

## 6.7 NEJČASTĚJŠÍ OMYLY PŘI INTEGRAČNÍCH SNAHÁCH

### 6.7.1 Jeden produkt pokryje všechny požadavky na IS

Je až zarážející, kolik manažerů integračních projektů hledá „vhodný integrační produkt“, který by jim vyřešil všechny problémy s využíváním informačních zdrojů. Je nutné si uvědomit, že žádné hotové softwarové řešení nemůže na 100% pokrýt všechny potřeby konkrétní instituce. Vždy bychom měli postupovat podle osvědčených metod projektování, kdy jednotlivé požadavky na IS popisujeme v podobě dílčích funkcí. K jejich zabezpečení potom můžeme hledat vhodné produkty, které lze různě kombinovat, případně dotvářet vlastním softwarovým vývojem.

**Požadavky na IS bychom měli rozdělit na jednotlivé funkční moduly a pro ně potom hledat softwarová řešení. Nebraňme se kombinacím produktů.**

### 6.7.2 Sjednotíme standardy a konfigurace všech dílčích informačních zdrojů

Ve většině případů usilujeme o integraci externích zdrojů, nad jejichž technologiemi nemáme přímou kontrolu – nemůžeme např. nutit provozovatele databází ProQuest, aby používali jiné nastavení Z39.50 nebo podporovali kódování CP 1250. Ale i v případě, že jsou provozovatelé informačních zdrojů ochotni přizpůsobit nastavení a standardizovat komunikační rozhraní, nelze jednoduše dosáhnout shody. Technologie, standardy a konfigurace totiž nestačí pouze jednorázově sjednotit na počátku projektu. Externí komponenty a zdroje se dynamicky vyvíjejí, a tak by bylo třeba jednotné chování stále udržovat. Různé komponenty se navíc vzájemně ovlivňují, vznikají složité vztahy jedinečné pro každý informační zdroj. Tento problém se projevil při zavádění protokolu Z39.50 v ČR – přestože byl vyhlášen společný Z39.50 profil, dodavatelé AKS i provozovatelé nadále používali různé způsoby indexování, takže Z39.50 zdroje ani s podporou profilu nevracely vždy srovnatelné výsledky.

**Musíme se smířit s různorodostí externích prvků a tu v integračním systému řešit.**

Poznámka: Výjimkou mohou být institucionální nebo regionální projekty, jejichž účastníci jsou řízeny společným orgánem, který může použité technologie ovlivnit nebo dokonce diktovat. Příkladem může být slovenský projekt KIS3G, který centralizoval AKS knihoven spadající pod Ministerstvo kultúry SR.

### 6.7.3 Všechny dokumenty budou v blízké době v elektronické podobě

Mnoho integračních projektů uvažuje pouze elektronické informační zdroje. Management těchto projektů často podléhá nesprávnému dojmu, že všechny cenné informace lze dnes nalézt v elektronických databázích a na internetu. Tento názor je zcela nesprávný. Klasická, zejm. knižní produkce, je díky moderním publikačním technologiím v nebývalém rozmachu a velký objem informací a znalostí je publikován právě tradičním tiskem. Autoři a vydavatelé chtějí publikováním generovat zisk, proto nechtějí publikovat na otevřeném internetu ani v elektronických databázích, kde ztrácejí přehled o využití. Autoři chtějí za své práce zaplatit, vydavatelé knih musí tyto peníze získat prodejem. Situace je odlišná pouze v oblasti elektronických časopisů, kde jsou příjmy zajištěny předplatným nebo prodejem jednotlivých článků. U monografií však stále žádný obchodní model elektronické distribuce dobře nefunguje.

**Ignorování klasického fondu v integračním systému je velkou chybou. Klasický fond je třeba propojit pomocí bibliografických databází a navazujících služeb EDD.**

### 6.7.4 Uživatelé nemusí být zapojeni do projektování

Projektoví manažeři a vývojáři se často domnívají, že přesně vědí, co uživatelé potřebují a co je pro uživatele nejlepší. Je to běžný omyl profesionálů, kteří se již dlouho pohybují v dané oblasti. Na projekt to však může mít velmi negativní dopady, protože cíle projektu neodpovídají potřebám uživatelů. Je třeba si uvědomit, že se uživatelské potřeby a hlavně uživatelské návyky díky rozvoji technologií rychle mění. Je proto třeba průběžně identifikovat potřeby uživatelů. Toho lze dosáhnout pouze zapojením všech typů uživatelů do projektování. Je třeba znát jejich očekávání a potřeby, je třeba je zapojit do připomínkování a testování. Testováním by měl projít systém jako celek, protože každá část ovlivňuje chování celku. Pokud není možné do testování



zapojit externí uživatele, je vhodné využít zkušeností pracovníků služeb a podpory, kteří přicházejí s externími uživateli nejvíce do styku a znají tak nejlépe jejich potřeby.

**Na vývoji IS by se měli podílet reprezentanti všech, kdo bude s IS pracovat: management, operační pracovníci, administrátoři a koncoví uživatelé.**

### **6.7.5 Finanční náročnost projektu je dána pořizovacími náklady**

Při sestavování rozpočtu projektu se často uvažuje pouze fáze přípravy a zprovoznění IS. To je kritická chyba, protože projekt je nutné finančně zajišťovat po celou dobu jeho životního cyklu, tedy až do ukončení provozu. Je proto nutné uvažovat souhrn všech dílčích nákladů, od nákupu software a hardware, přes zajištění mezd pracovníků, jejich proškolení, hrazení ročních údržbových poplatků a poplatků za obnovování licencí až po náklady za propagaci. Tento souhrn tvoří celkové náklady na vlastnictví - náklady, které musíme být schopni finančně pokrýt. Dejme si také pozor na obchodní nabídky dodavatelů, kteří často vytvářejí kalkulace tak, aby se na první pohled jevily co nejpříznivěji, ale přitom obsahují mnoho skrytých nákladů, které se do celkového součtu v nabídce nepromítají.

**Prvotní pořizovací náklady mohou být oproti celkovým nákladům na vlastnictví zanedbatelné.**

### **6.7.6 Při spuštění IS musí být realizovány všechny plánované funkce**

Manažeři projektů pro integraci informačních zdrojů často usilují o to, aby byly v IS v okamžiku spuštění zprovozněny všechny plánované funkce a připojeny všechny uvažované informační zdroje. Takový stav by byl samozřejmě ideální, nicméně u složitých projektů to není z časových důvodů možné. Vývoj IS je časově náročný a finančně nákladný. Mnoho projektů skončilo nezdarem právě proto, že několik let probíhaly přípravy a vývoj, a aniž by byl jediný modul zprovozněn a nabídnut uživatelům, objevil se podobný konkurenční projekt, který oslovil potencionální uživatele. O projektovaný IS už najednou nebyl zájem a vývoj byl ukončen. Příkladem takového nezdaru může být např. vývoj českého AKS Relief.

**IS by měl být spuštěn v krátkodobém časovém horizontu se základní nabídkou služeb, která by ale neměla být příliš úzká, aby neodradila uživatele. Postupně by pak měly být přidávány další funkce či moduly a celý IS zdokonalován.**

### **6.7.7 Není třeba nasazovat systém řízení jakosti**

Manažeři projektů bývají přesvědčeni, že mají vedení projektu pevně v rukou a mohou tak bez systému řízení jakosti pohlídat kvalitu služeb i celého IS. QMS systémy však zajišťují pravidelný monitoring všech klíčových funkcí, kterých mohou být ve složitějších projektech stovky. Není v lidských silách bez jakýchkoli pomůcek QMS evidovat a efektivně reagovat na všechny odezvy uživatelů, kontrolovat události v systémových reportech, analyzovat případné výpadky informačních zdrojů a dílčích komponent, hodnotit chování IS v nových verzích OS či prohlížečů, monitorovat a hodnotit bezpečnost rizika apod.

**Nasazení systému řízení jakosti je dnes kvůli složitosti a komplexnosti informačních systémů pro integraci informačních zdrojů nezbytností.**

### **6.7.8 S integračním systémem začnou uživatelé sami pracovat**

Je až zarážející, jak vysoké finanční prostředky se vynakládají na technickou realizaci informačních systémů a jak malá pozornost i finance jsou věnovány na jejich propagaci a podporu. Uživatelé často vůbec netuší, že mohou služeb IS využívat a jaký užitek jim mohou přinést. Není jim poskytována možnost proškolení a téměř žádná podpora. Příčiny nízké návštěvnosti IS se potom chybně hledají v nekvalitních službách či nezájmu uživatelů.

**I sebelepší IS je třeba propagovat. Uživatelé se musí dozvědět o jeho existenci a hlavně pochopit, jaký přínos pro ně představuje. Pro koncové uživatele musí být poskytována kvalitní technická podpora, konzultace a školení či online kurzy.**

## 7 ZÁVĚRY

### 7.1 DOSAŽENÉ VÝSLEDKY

V oboru české informační vědy doposud chyběla podrobnější analýza architektury a technologických možností integrace informačních zdrojů, která se v mnoha ohledech liší od principů obecné systémové integrace. Následný benchmarking jednotlivých vlastností IS pro integraci informačních zdrojů a nejlepší doporučení pro projektování těchto systémů dosud v této oblasti zcela chyběl. Rozhodl jsem se proto tuto oblast pečlivě prozkoumat a definovat, analyzovat, stanovit a vyložit nejlepší zkušenosti a doporučení (tzv. best practices) pro budování informačních systémů pro integraci informačních zdrojů v oblasti knihovních služeb.

Výsledkem analýz realizovaných integračních projektů a současných technologických možností v oblasti integrace informačních zdrojů je soupis nejlepších postupů a doporučení pro řízení budoucích projektů. Všechna doporučení prošla přísnou revizí, aby byla v souladu s používanými standardy a znalostmi informační vědy. Byla také uzpůsobena pro prostředí českých knihoven, aby bylo možné doporučení aplikovat ve všeobecné praxi.

Byla použita ta doporučení, která jsou obecně platná a nejsou specifická pouze pro jeden konkrétní projekt. Pro projekty integrace informačních zdrojů je typická vysoká variabilita prvků a použitých řešení, mnoho specifik proto nebylo možné zevšeobecnit. Důraz byl kladen na technologie a řešení, která jsou zaštitěna průmyslovými standardy a jsou šířeji zastoupena v realizovaných projektech.

Dalším výzkumem by měly být v budoucnu zhodnoceny metody distribuovaného vyhledávání založené na obecných standardech HTTP a XML, zejm. řešení SRU/SRW. Nejlepší postupy a doporučení pro implementaci SRU/SRW by mohlo výrazně přispět k většímu rozšíření této nadějně technologie mezi české knihovny a provozovatele informačních zdrojů.

### 7.2 POUŽITÉ METODY

V procesu benchmarkingu (hledání nejlepších postupů a doporučení) bylo v dizertační práci použito několik metod, které bylo nutné volit a kombinovat v závislosti na povaze a možných metrikách jednotlivých vlastností IS. Konkrétně se jednalo o tyto metody:

- porovnávání cílů a úkolů v projektové dokumentaci se závěrečnými zprávami a skutečným stavem projektů
- simulací auditu, kterým byly posouzeny cíle s ohledem na stanovené parametry kvality, finančních nákladů, rozsahu funkcí, výstupu funkcí apod.
- konzultace s manažery jednotlivých projektů
- analýza dotazů na podporu
- analýza systémových logů a reportů
- porovnávání vyhledávacích možností a relevance výsledků vyhledávání v nativních rozhraní informačních zdrojů s rozhraním použitého integračního systému

Na mnoho vlastností IS nebylo možné použít kvantifikační metrické metody. Pro každou posuzovanou vlastnost či proces IS bylo přiřazeno více metrik (indikátorů a ukazatelů), aby byla představa o míře shody co nejpřesnější.

### 7.3 VLASTNÍ VÝZKUM

V první fázi výzkumu jsem mapoval současnou situaci v oblasti integrace informačních zdrojů v ČR a zahraničí. Definoval jsem specifika integrace informačních zdrojů v porovnání s obecnou systémovou integrací. Zvláštní pozornost jsem věnoval studiu rizik integrace informačních zdrojů: tedy co integrací můžeme ztratit oproti využívání informačních zdrojů jednotlivě.

Informační zdroje, se kterými knihovny pracují a staví na nich své služby, jsem následně rozdělil do tří základních skupin, které se možnostmi integrace výrazně liší. Jedná se o klasický fond knihoven, o elektronické databáze a archivy a o volný prostor internetu (mělký i hluboký web).

**Klasický fond knihoven** je tvořen jednotkami v analogové či elektronické formě, které jsou fixovány na fyzická přenositelná media. Jsou to tedy všechny typy dokumentů v offline režimu. Protože se jedná o fyzicky uložený fond, nelze ho integrovat přímo, ale prostřednictvím elektronické evidence (OPAC) a pomocí zprostředkovatelských služeb, např. elektronického dodávání dokumentů. Vzácné části fondu se postupně digitalizují, zejm. historické a jedinečné sbírky, a zpřístupňují online. Knihovny mají se zpracováním klasického fondu velké zkušenosti a díky vysoké standardizaci zpracování (MARC/AACR2) dosahují i velmi dobrých výsledků ve vyhledávání a v procesu půjčování. Na zavedených standardech byly postaveny i poměrně úspěšné integrační projekty, zejm. ty, které využívají centralizovaných modelů (fyzické souborné katalogy, soubory národních autorit apod.). S rozšířením technologie Z39.50 se objevily také projekty založené na distribuovaném modelu (virtuální souborné katalogy, informační brány apod.). Oblastí, kterou se prozatím nedaří příliš rozvíjet, jsou navazující služby, které zajišťují dodávání jednotek klasického fondu uživatelům elektronickou cestou. Tím jsou integrační systémy výrazně ochuzeny, protože uživatelům nenabízejí primární informační objekty. Podrobně jsem analyzoval příčiny stagnace těchto služeb. Ukázalo se, že hlavní vinu na tom nesou v ČR nedokonalé mikroplatební systémy (vysoká cena za transakci, nedůvěra uživatelů, nutnost skládat kredit), komplikace a omezení daná autorským zákonem a licenčními smlouvami, nedostatek pracovních sil na tyto služby (integrační systémy s těmito činnostmi v projektech často vůbec nepočítají) a částečně i absence centrálního národního systému elektronického dodávání dokumentů, který by nabízel fond všech zúčastněných knihoven a zároveň by nabízel jednotnou databázi uživatelů a společný platební systém. Záměrně nebyla v této práci analyzována digitalizace klasického fondu, protože tato problematika spadá do zcela samostatné oblasti a nebylo by možné ji dostatečně podrobně prozkoumat.

**Elektronické online databáze a archivy** představují vlastní nebo zakoupené digitální sbírky a databáze, které tvoří elektronický knihovní fond dostupný uživatelům pouze online. Řada knihoven vytváří cenné faktografické databáze nebo sbírky zdigitalizovaných starých tisků, rukopisů či historických časopisů. Kupují také licence plnotextových nebo bibliografických databází komerčních vydavatelů a elektronické knihy (eBooky). Pro tyto informační zdroje je typické, že informace v nich obsažené jsou alespoň v minimální míře organizovány a popsány. Databáze a archivy s primárními informačními objekty jsou často vybaveny i sekundárními informacemi, na kterých mohou být vybudovány vyhledávací indexy nebo které mohou být nabízeny k externímu využití (např. sklizení). Tato skupina zdrojů je tedy na integraci nejlépe připravena, i když i zde se setkáváme s řadou dílčích problémů, jako jsou např. různé způsoby indexování, různé vyhledávací body, nekompatibilní komunikační rozhraní apod.

Téměř bezbřehou oblast informačních zdrojů představuje tzv. **mělký web**, tedy volný prostor internetu tvořený statickými dokumenty a statickými sídly, které mají pevnou adresu URL. Mohou to být dokumenty a soubory nejrůznějších formátů. Tato část internetových zdrojů je relativně dobře zmapovatelná běžnými indexovacími a vyhledávacími nástroji, jejichž technologie jsou založeny na indexování pomocí sledování odkazů URL. Informační zdroje spadající do této skupiny jsou díky pevné adrese URL „uchopitelné“. Z hlediska využití a integrace je nejproblematičtější, ale informačně velice cenná, oblast, kterou tvoří dynamicky generované dokumenty a prohledatelné databáze na internetu. Označuje se často jako neviditelný nebo **hluboký web**. Informační zdroje této skupiny nelze většinou automaticky indexovat a prohledávat, takže se uživatelé nemohou o obsahu těchto zdrojů dovědět pomocí běžných vyhledávačů. Pro přístup k dynamickému obsahu je třeba zadávat vstupní data nebo použít vyhledávací jazyk rozhraní. Informační zdroje databázového typu také často vyžadují autentikaci, kdy je pro vstup do systému nutné zadat ID a heslo.

**Knihovny mohou v této oblasti nabídnout nenahraditelné služby s vysokou přidanou hodnotou.** Jednou z těchto hodnot je precizní a přehledný popis pečlivě vybraných internetových zdrojů. Uživatelé tak mohou k jednotlivým vědním oborům získat seznam prověřených a kvalitních informačních zdrojů, které by jinak velmi těžko hledali pomocí internetových vyhledávačů či internetových katalogů.

---

Následující část práce podrobně analyzuje, jakým způsobem je možné informační zdroje integrovat s ohledem na bezpečnost, autentizační mechanizmy, popis informačních zdrojů a zejm. vyhledávání a následnou prezentaci vrácených dat. Porovnává distribuované a centralizované vyhledávání, paralelní a sekvenční vyhledávání, vyhledávací procesy podle míry strukturovanosti a míry digitalizace informačních zdrojů, vícejazyčnosti apod.

Významnou částí integračních systémů jsou navazující služby, a proto jim byla věnována samostatná analytická část práce.

Před závěrečnou, stěžejní kapitolou byly ještě vloženy části, zabývající se nejběžnějšími integračními technologiemi současnosti, poznámky o zajištění kvality a modelový příklad integračního projektu. Cílem těchto kapitol bylo doplnit informace nutné ke správné interpretaci kapitoly nejlepšími postupy a doporučeními.

Zatímco kapitoly 1 až 5 byly pojaty analyticky, závěrečná 6. kapitola shrnuje typické problémové oblasti projektování integrace informačních zdrojů a formou seznamu krátkých a věcných poznámek definuje nejlepší postupy a doporučení.

## 8 SLOVNÍK ZKRATEK

### **AES**

*Advanced Encryption Standard*

Druh symetrické šifrovací technologie, kterou americký normalizační úřad NIST vybral jako národní standard pro šifrování obchodní a vládní komunikace. Standard AES je postaven na šifrovacím algoritmu Rijndael. AES podporuje řada produktů, včetně SSL serverů.

### **CCL**

*Common Command Language*

Dotazovací jazyk pro prohledávání online databází schválený jako standard ISO. CCL se využívá např. v protokolu Z39.50.

### **COUNTER**

*Counting Online Usage of Networked Electronic Resources*

Mezinárodní iniciativa pro zvýšení spolehlivosti statistických ukazatelů využívání online informačních zdrojů a služeb.

### **DBMS**

*Database Management System*

Systém řízení báze dat. Softwarová aplikace, která zajišťuje práci s úložištěm dat a která tak vytváří rozhraní mezi aplikačními programy a uloženými daty.

### **DOI**

*Digital Object Identifier*

Identifikátor digitálních objektů. Prostředek k trvalé a jednoznačné identifikaci objektu intelektuálního vlastnictví (díla) v síťovém prostředí, nezávislý na aktuálním umístění objektu.

### **DOM**

*Document Object Model*

Objektový model dokumentu, který hierarchicky reprezentuje XML nebo HTML dokument. DOM umožňuje přístup či modifikaci obsahu, struktury nebo stylu dokumentu. Typickým využitím DOM jsou aktivní skripty (např. v JavaScriptu) na straně webového klienta. Významným rysem DOM je jazyková i platformová nezávislost.

### **DRM**

*Digital Rights Management*

Systémy pro řízení digitálních práv. Zajišťují ochranu autorských práv objektů, které jsou distribuovány prostřednictvím internetu nebo jiných médií. Většinou jsou založeny na šifrování obsahu, ke kterému lze přistoupit pouze na základě autentikace.

### **EDD**

*Electronic Document Delivery; Elektronické dodávání dokumentů*

Služba zajišťující dodávání elektronických kopií článků či jiných dokumentů z fondů knihovny koncovému uživateli. Dokumenty jsou naskenovány a ve formě kvalitních snímků (typicky ve formátu PDF nebo JPG) dodány prostřednictvím e-mailu nebo WWW na pracovní stanici uživatele.

### **ERM**

*Electronic Resource Management*

Systém pro správu elektronických informačních zdrojů. Jeho hlavní funkcí je zajištění podpory všech fází životního cyklu elektronických zdrojů a správa licencí. Často také slouží jako webový katalog elektronických zdrojů pro uživatele.

### **EXIF**

*Exchangeable image file format*

Formát pro ukládání metadat do obrazových souborů, zejm. s kompresí JPEG. Používaný zejm. pro ukládání informací o nastavením digitálního fotoaparátu v čase expozice daného snímku.

**FTP***File Transfer Protocol*

Protokol určený pro přenos souborů mezi dvěma počítači, zejm. mezi klientem a serverem.

**Google Bomb**

Snaha ovlivnit pořadí daného dokumentu v seznamu výsledků vráceného vyhledávacím strojem Google, často s komerčním, politickým nebo žertovným úmyslem.

**HTML***Hypertext Markup Language*

Značkovací jazyk, který používá definované značky k vytváření a formátování dokumentů, zejm. webových stránek. Jeho definici schvaluje konsorcium W3C. Jazyk HTML je postupně vytlačován novým značkovacím jazykem XHTML, který vyhovuje syntaxi XML.

**HTTP***Hypertext Transfer Protocol*

Stavový protokol používaný mezi klientem a serverem pro přenos dokumentů na internetu. Protokol HTTP byl původně určen pro přenos HTML dokumentů, ale dnes je jeho využití daleko širší.

**HTTPS***Hypertext Transfer Protocol over SSL*

Zabezpečená verze protokolu HTTP, která znesnadňuje odposlechnutí komunikace.

**ID3**

Formát pro ukládání metadat do zvukového souboru MP3. Může obsahovat údaje o názvu skladby, skladateli, interpretovi, hudebním žánru, roku vydání atd.

**IM***Instant messaging*

Forma elektronické komunikace, která umožňuje interakci mezi dvěma či více uživateli, kteří jsou v danou chvíli online. Podobá se telefonické konverzaci, probíhá však v textové formě. Nejrozšířenějšími systémy IM jsou ICQ, AIM, Skype a MSN.

**Inteligentní mobilní telefon***smartphone*

Kombinace mobilního telefonu a PDA. Inteligentní mobilní telefon (označovaný také jako chytrý mobil) se od klasických mobilních telefonů liší tím, že mají otevřený operační systém a úložiště dat, takže může uživatel do telefonu ukládat informace a instalovat další aplikace.

**IPTC***International Press Telecommunications Council*

Nesprávné, ale rozšířené pojmenování formátu IIM (Information Interchanges Model) pro ukládání metadat do obrazových souborů. Mohou obsahovat údaje o autorovi, datu a času vytvoření, klíčová slova, kategorie, urgence, kontakt, copyright apod. Formát je podporován např. v editoru Adobe Photoshop. Dnešní podoba formátu má již strukturu XML.

**ISAN***International Standard Audiovisual Number*

Mezinárodní standard pro jednoznačnou identifikaci audiovizuálních prací. Identifikátor tvoří 24 číslic a dokáže identifikovat nejen vlastní dílo, ale i epizodu/část a verzi.

**ISBN***International Standard Book Number*

Mezinárodní standard ISO 2108 pro jednoznačnou identifikaci knižních publikací. Existuje ve verzi ISBN-10 a ISBN-13 podle počtu znaků tvořících kód.

**ISMN***International Standard Music Number*

Mezinárodní standard ISO 10957, představuje desetipísmenný alfanumerický identifikátor tištěných hudebnin.

**ISRC***International Standard Recording Code*

Mezinárodní standard ISO 3901 pro jednoznačnou identifikaci zvukových nahrávek a hudebních video nahrávek.

**ISRN***International Standard Technical Report Number*

Mezinárodní standard ISO 10444 pro jednoznačnou identifikaci technické dokumentace.

**ISSN***International Standard Serial Number*

Mezinárodní standard ISO 3297 pro jednoznačnou identifikaci tištěných a elektronických periodik. Kód je tvořen dvojicí čtyřech číslic.

**JPEG***Joint Photographic Experts Group*

Široce rozšířený grafický formát pro ztrátovou komprimaci bitmapových obrazových dat. Formát dosahuje vysoce účinné komprese až 5% původní velikosti, avšak za cenu ztráty kvality. JPEG podporují všechny webové prohlížeče i většina digitálních zařízení, jako jsou fotoaparáty a skenery.

**LDAP***Lightweight Directory Access Protocol*

Protokol pro přístup k adresářovým službám internetu nebo lokálních sítí. Umožňuje definovat organizační strukturu a vyhledávat v ní organizační jednotky, osoby a technické prostředky. Slouží pro účely autentikace a získávání informací o ověřovaných objektech. Je to odlehčená verze standardu X.500 pro adresářové služby v síti.

**MARC***MAchine-Readable Cataloging*

Datový formát vyvinutý Knihovnou amerického Kongresu pro účely výměny bibliografických informací. Postupně se MARC stal interním datovým formátem většiny knihovních systémů a bibliografických databází. MARC má mnoho národních modifikací, v poslední době však sílí snaha všechny formy sjednotit do podoby MARC 21. Objevují se snahy postavit MARC na XML, aby mohl lépe spolupracovat s moderními IS, zejm. digitálními sbírkami.

**METS***Metadata Encoding and Transmission Standard*

Mezinárodní standard využívající prostředky XML pro kódování popisných, administrativních a strukturálních metadat objektů digitálních knihoven. Standard METS spravuje agentura Network Development and MARC Standards Office při Knihovně amerického Kongresu v rámci iniciativy Federace digitálních knihoven (Digital Library Federation).

**MP3**

Formát pro vysokou kompresi digitálních zvukových souborů. Vyvinutý pro potřeby přenosu zvukových souborů po internetu a k jejich ukládání v kapesních audiopřehrávačích a na zvukových serverech.

**OAI-PMH***Open Archives Initiative Protocol for Metadata Harvesting*

Protokol pro metadatovou interoperabilitu umožňující poskytovatelům služeb automatizované získávání (sklizení) metadat nabízených poskytovateli dat z různých úložišť (repozitářů).

**OpenURL**

Standard pro přenos metadatových (zejm. bibliografických) informací pomocí URL mezi knihovnickými aplikacemi. OpenURL využívají zejm. linkovací servery, jako je SFX. OpenURL byl schválen jako ANSI standard Z39.88.

**OS***Operation system; Operační systém*

Software, který řídí a spravuje hardware a zajišťuje základní systémové operace. Vytváří systémové prostředí pro spouštění programových aplikací.

**PDF***Portable Document Format*

Dokumentový formát pro bezpečnou a spolehlivou distribuci dokumentů. Formát zachovává písma, grafické objekty a rozložení zdrojového dokumentu v kompaktní formě nezávisle na aplikaci a platformě zdrojového a cílového počítače. Pro čtení je třeba použít speciální prohlížeč, typicky Adobe Reader.

**PKCS***Public Key Cryptography Standards*

Řada standardů pro kryptografii s veřejným klíčem, která zahrnuje šifrování RSA, shodu klíčů Diffie-Hellman, šifrování založené na heslech, rozšířenou syntaxi, syntaxi kryptografických zpráv, syntaxi informací soukromých klíčů a syntaxi žádostí o certifikát, včetně vybraných atributů. Tyto standardy byly vyvinuty společností RSA Data Security, Inc., která je spravuje a je jejich vlastníkem.

**PNG***Portable Network Graphics*

Volný grafický formát vyvinutý jako nástupce patentově chráněného formátu GIF. Nabízí až o 30% lepší kompresi než GIF, podporuje vkládání textových informací a průhlednost. PNG podporují všechny webové prohlížeče.

**QMS***Quality Management System*

Systém pro řízení jakosti, používaný zejm. v souvislosti s normami ISO 9001. Souhrn postupů a opatření organizace směřujících k dosažení certifikované jakosti.

**RDF***Resource Description Framework*

Univerzální, sémanticky citlivý jazyk pro popis zdrojů. Každý zdroj je identifikován pomocí zcela jednoznačného identifikátoru URI (Uniform Resource Identifier). Zdroje jsou popisovány pomocí vlastností a jejich hodnot. RDF je často prezentován formou XML. Definici jmenného prostoru RDF definuje standard RDFS (RDF schema), jakým může být např. Dublin Core. Standard RDF je využíván např. v technologii RSS, mohl by se stát také prostředkem pro budování sémantického webu.

**RSS***Really Simple Syndication*

Formát pro publikování informací na webu, zejm. v IS s rychle se měnícím obsahem. RSS dokument je vytvářen v syntaxi XML a obsahuje krátký popis (typicky název a anotaci) doplněný o odkaz na plnou verzi daného objektu. Uživatelé pak generované RSS zprávy odebírají prostřednictvím tzv. RSS kanálů do svých RSS čteček.

**SAML***Security Assertion Markup Language*

Značkový jazyk určený k výměně autentizačních a autorizačních údajů mezi dvěma zabezpečenými systémy, typicky mezi poskytovatelem identity a poskytovatelem konkrétní služby. Standard SAML vyvíjí sdružení OASIS Security Services Technical Committee.

**SDI***Selective Dissemination of Information*

Adresné šíření informací. Systém pro rozesílání informací o nejnovějších přírůstcích na základě uživatelem definovaného dotazu, který je v pravidelných intervalech automaticky spouštěn.

**SICI***Serial Item and Contribution Identifier*

Mezinárodní standard ANSI/NISO Z39.56 pro jednoznačnou identifikaci ročníků, článků a dalších částí periodik.

**SOAP***Simple Object Access Protocol*

Protokol založený na XML, který umožňuje komunikaci jakéhokoli objektu v distribuovaném prostředí. Protokol SOAP se používá při vývoji webových služeb.



**SSL***Secure Sockets Layer*

Protokol pro zabezpečení datového přenosu na internetu. SSL používá k šifrování systému privátního a veřejného klíče.

**SSO***Single sign-on*

Technologie umožňující využití jediného procesu autentikace a autorizace k přístupu uživatele do všech IS a zařízení, ke kterým má uživatel povolen přístup, bez nutnosti zadávat přihlašovací údaje opakovaně.

**URI***Uniform Resource Identifier*

Jednoznačný identifikátor informačního objektu na internetu. Označovaný také jako „adresa“.

**URL***Uniform Resource Locator*

Typ URI, který označuje umístění informačního objektu na internetu a zároveň k němu umožňuje přístup. Typicky URL určuje komunikační protokol přístupu k objektu, název hostitelského serveru a cestu k objektu, např.: <http://www.cuni.cz/historie/listina.html>

**URN***Uniform Resource Name*

Typ URI, který slouží jako trvalý, na umístění nezávislý identifikátor informačního objektu na internetu.

**VPN***Virtual Private Network*

Prostředek k propojení několika počítačů na různých místech internetu mezi sebou bezpečným způsobem. Mezi serverem, který VPN zajišťuje, a klienty vznikají síťové tunely, v rámci nichž lze navazovat další síťová spojení. Typicky se VPN používá k připojení poboček a domácích pracovišť do lokální sítě organizace.

**WAP***Wireless Application Protocol*

Sada komunikačních protokolů, které definují způsob, jak bezdrátová zařízení (jako jsou mobilní telefony a PDA) přistupují k internetovým online službám, např. k elektronické poště či WWW.

**WSDL***Web Service Description Language*

Standardní formát pro popis webových služeb. Popisuje, jak k dané webové službě přistupovat a jaké operace budou prováděny.

**XML***Extensible Markup Language*

Značkovací jazyk, který popisuje strukturu dat. Není to sada standardizovaných značek jako HTML, ale metajazyk, podobně jako SGML. XML umožňuje definovat vývojářům vlastní značky.

**Z39.50**

Protokol, který slouží pro vyhledávání a získávání informací ze vzdálených online databází nezávisle na jejich nativním rozhraní.

## 9 LITERATURA

1. AMBROŽ, Jan. *Mikroplatby zamrzly. Nahradí je SMS? (2.)* [online]. 2005 , 13.04.2005 [cit. 2007-02-05]. Dostupný z WWW: <[http://www.ambroz.org/136142\\_clanek.php](http://www.ambroz.org/136142_clanek.php)>.
2. Amos software. *Adobe LiveCycle Policy Server* [online]. 2004 [cit. 2005-04-10]. Dostupný z WWW: <<http://www.amsoft.cz/Produkty/Adobe/server/policy/overview.html>>.
3. BLESSINGER, Kelly, OLLE, Maureen. Content analysis of the leading general academic databases. *Library, Collections, Acquisitions & Technical Services* [online]. 2004, no. 28 [cit. 2007-03-21], s. 335-346. Dostupný z WWW: <[projects.ics.hawaii.edu/~jacso/PDFs/blessinger-content-analysis-leading-general.pdf](http://projects.ics.hawaii.edu/~jacso/PDFs/blessinger-content-analysis-leading-general.pdf)>.
4. BOHÁČEK, Martin. *Stručný komentář autorského zákona pro SKIP a knihovny* [online]. 1. varianta. Praha : Národní knihovna ČR, 2001 [cit. 2005-04-08]. Dostupný z WWW: <[http://www.nkp.cz/o\\_knihovnach/AutZak/AZkucharka.htm](http://www.nkp.cz/o_knihovnach/AutZak/AZkucharka.htm)>.
5. BOYD, John, et al. The One-Box Challenge : Providing a Federated Search That Benefits the Research Process. *Serials Review* [online]. 2006, vol. 32, is. 4 [cit. 2007-07-28], s. 247-254. Dostupný z WWW: <<http://www.sciencedirect.com/science/article/B6W63-4M0J4M6-1/2/932bddf79cc80eeeea14ab367907ecc7>>.
6. BrightPlanet Corporation. *Six Major Trends Affecting Knowledge Management and Information Technology*. Washington : BrightPlanet Corporation, 2004. 9 s.
7. CORTES, Leslie. *Designing a Graphical User Interface* [online]. *Medical Computing Today*, May 1997 [cit. 2005-01-27]. Text v angličtině. Dostupný z WWW: <<http://www.medicalcomputingtoday.com/0agui.html>>.
8. CRAWFORD, Walt. Library 2.0 and "Library 2.0". *Cites & Insights : Crawford at Large* [online]. 2006, vol. 6, no. 2 [cit. 2007-07-28], s. 1-32. Dostupný z WWW: <<http://citesandinsights.info/civ6i2.pdf>>. ISSN 1534-0937.
9. Často kladené otázky k novým normám ISO 9001:2000. Z angl. orig. přeložil a doplnil Jiří Kruliš. 1. vyd. Praha : Český normalizační institut, 2001. 26 s. Aktuálně o ISO 9000:2000. ISO Frequently Asked Questions. ISBN 80-7283-063-5.
10. ČSN ISO 690 *Dokumentace. Bibliografické citace : obsah, forma, struktura*. Praha : Český normalizační institut, 1996. 31 s.
11. ČSN ISO 690-2 *Bibliografické citace. Část 2: Elektronické dokumenty nebo jejich části*. Praha : Český normalizační institut, 2000. 24 s.
12. DEMPSEY, Lorcan. The recombinant library: portals and people. *Journal of Library Administration* [online]. 2003, vol. 4, no. 39 [cit. 2007-02-07], s. 103-136. Dostupný z WWW: <[http://www.oclc.org/research/staff/dempsey/recombinant\\_library/dempsey\\_recombinant\\_library.htm](http://www.oclc.org/research/staff/dempsey/recombinant_library/dempsey_recombinant_library.htm)>. ISSN 0193-0826.
13. Digimarc. *MyPictureMarc : Communicate your copyrights with MyPictureMarc* [online]. 2004 [cit. 2005-04-10]. Dostupný z WWW: <<http://www.digimarc.com/watermark/mypicturemarc/>>.
14. DOSEDĚL, Tomáš. *Počítačová bezpečnost a ochrana dat*. 1. vyd. Brno : Computer Press, 2004. 182 s. ISBN 80-251-0106-1.
15. *E-government in New Zealand : New Zealand E-government Programme* [online]. 2001-2007 [cit. 2007-05-20]. Dostupný z WWW: <<http://www.e.govt.nz/>>.
16. HANÁČEK, Petr, STAUDEK, Jan. Správa identity. In HRUŠKA, Tomáš. *DATAKON 2005 : Brno, 22.-25. 10. 2005*. [s.l.] : [s.n.], 2005. s. 123-146. Dostupný z WWW: <[www.fi.muni.cz/usr/staudek/vyuka/security/d05\\_idm\\_tutorial\\_text.pdf](http://www.fi.muni.cz/usr/staudek/vyuka/security/d05_idm_tutorial_text.pdf)>. ISBN 80-210-3813-6.

17. Juniper Networks. *Význam standardů pro řízení síťového provozu*. [s.l.] : [s.n.], 2006. 10 s. White Paper.
18. KANIGEL, Robert. *The one best way : Frederick Winslow Taylor and the enigma of efficiency*. Cambridge : MIT Press, 2005. 675 s. ISBN 0-262-61206-2.
19. KAŇKOVSKÝ, Pavel. Bezpečnostní slabiny WWW aplikací. In HRUŠKA, Tomáš. *DATAKON 2005 : Proceedings of the Annual Database Conference*. 1. vyd. Brno : Masarykova univerzita, 2005. s. 41-55. ISBN 80-210-3813-6.
20. KATUŠČÁK, Dušan. *KIS3G : Knižnično-informačný systém tretej generácie*. Martin : Slovenská národná knižnica, 2005. 11 s.
21. KAY, Russell. *Authentication : QuickStudy* [online]. Framingham : Computerworld, 2000 [cit. 2005-04-11]. Dostupný z WWW: <<http://www.computerworld.com/q?a5630>>.
22. KAY, Russell. *Biometric authentication : QuickStudy* [online]. Framingham : Computerworld, 2005 [cit. 2005-04-11]. Dostupný z WWW: <<http://www.computerworld.com/q?53319>>.
23. KOUŘIL, Daniel. Bezpečnost v distribuovaném prostředí. *Zpravodaj ÚVT MU : Bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě*. 2005, roč. 15, č. 4, s. 2-6. Dostupný z WWW: <<http://www.ics.muni.cz/bulletin/issues/vol15num04/kouril/kouril.html>>.
24. KREJČÍ, Richard. *Elektronický podpis v PDF* [online]. Praha : Grafika Publishing, 2002 [cit. 2005-04-10]. Dostupný z WWW: <<http://www.grafika.cz/art/pdf/pdfpodmis.html>>.
25. KUBA, Martin. Web Services. In VOJTÁŠ, Peter, SKOPAL, Tomáš. *DATAKON 2006 : Proceedings of the Annual Database Conference*. 1. vyd. Brno : Masarykova univerzita, 2006. s. 93-112. ISBN 80-210-4102-1.
26. LARGE, Andy, TEDD, Lucy. *Automatizace knihovnických a informačních služeb*. 1. vyd. Budapest : Open Society Institute, 1996. 85 s. Regional Library Program.
27. Library of Congress. *Preserving Our Digital Heritage : Plan for the National Digital Information Infrastructure and Preservation Program* [online]. c2002 [cit. 2005-04-29]. Dostupný z WWW: <[http://www.digitalpreservation.gov/repor/ndiipp\\_plan.pdf](http://www.digitalpreservation.gov/repor/ndiipp_plan.pdf)>.
28. LOSHIN, Pete. *Single Sign-on : QuickStudy* [online]. Framingham : Computerworld, 2001 [cit. 2005-04-11]. Dostupný z WWW: <<http://www.computerworld.com/q?a5640>>.
29. MALÝ, Jaroslav. *Projektování informačních systémů I*. 1. vyd. Hradec Králové : Gaudeamus , 1998. 147 s. ISBN 80-7041-518-5.
30. MARTINKOVÁ, Alice. Uživatelská rozhraní knihovnicko-informačních systémů pro těžce zrakově postižené uživatele. In *Knihovny současnosti 2006*. 1. vyd. Brno : Sdružení knihoven ČR, 2006. s. 210-224. ISBN 80-8649-41-7.
31. Microsoft ČR. *Integrace podnikových aplikací* [online]. [2004] [cit. 2005-01-25]. Dostupný z WWW: <<http://www.microsoft.com/cze/enterprise/eai/>>.
32. Národní knihovna ČR. Česká terminologická databáze knihovnictví a informační vědy [online]. 2001-2002 [cit. 2005-01-21]. Dostupný z WWW: <[http://sigma.nkp.cz/F/?func=file&file\\_name=find-a&local\\_base=kt](http://sigma.nkp.cz/F/?func=file&file_name=find-a&local_base=kt)>.
33. NEEDLEMAN, Mark. Z39.50 - A Review, Analysis and Thoughts on the Future. *Library Hi Tech*, 2000, vol. 18, no. 2, s. 158-169. ISSN 0737-8831.
34. NEMEŠKALOVÁ, Hana. Srovnání funkcionality MetaLib/SFX v Jednotné informační bráně se zahraničními projekty. *Knihovna plus* [online]. 2005, č. 2 [cit. 2007-03-18]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovna52/nemeskalova2.htm>>. ISSN 1801-5948.
35. PAVLÍK, Jiří. Shibboleth - elegantní technologie pro vzdálený přístup k databázím. In *INFORUM 2006: 12. konference o profesionálních informačních zdrojích : Praha, 23. - 25.5. 2006*. [s.l.] : [s.n.], 2006. s. 3. Dostupný z WWW: <[http://www.inforum.cz/inforum2006/pdf/Pavlik\\_Jiri.pdf](http://www.inforum.cz/inforum2006/pdf/Pavlik_Jiri.pdf)>.

36. POKORNÝ, Jan. Osvobodte své CD-ROMy!. *U nás : Knihovnicko-informační zpravodaj*. 2000, roč. 10, č. 1, s. 19-21. Dostupný z WWW: <[http://unas.svkhk.cz/pdf\\_archiv/0345.pdf](http://unas.svkhk.cz/pdf_archiv/0345.pdf)>.
37. ROSA, Zdeněk, et al. Uplatnění požadavků normy ISO 9001:2000 v praxi. 1. vyd. Praha : Český normalizační institut, 2001. 116 s. ISBN 80-7283-051-1.
38. ROSENAU, Milton D. Řízení projektů. Eva Brumovská. 1. vyd. Praha : Computer Press, 2000. 344 s. Succesfull Project Management. ISBN 80-7226-218-1.
39. ROSENMAYER, Jiří. Bezpečnostní audit Firewall a IDS. In VOJTÁŠ, Peter, SKOPAL, Tomáš. DATAKON 2006 : Proceedings of the Annual Database Conference. 1. vyd. Brno : Masarykova univerzita, 2006. s. 283-290. ISBN 80-210-4102-1.
40. SHEPHERD, Peter. COUNTER : towards reliable vendor usage statistics. *VINE*. 2004, vol. 34, no. 4, s. 184-189.
41. *Starting Out with Portals and OpenURL : An Introduction* [online]. Library of Congress, 2004 [cit. 2007-02-06]. Dostupný z WWW: <<http://www.loc.gov/catdir/lcpaig/introduction.html>>.
42. State Services Commision. *New Zealand Government Portal Strategy*. Wellington : State Services Commission, 2001. 46 s.
43. STOKLASOVÁ, Bohdana, KRBEČ, Pavel. *Jednotná informační brána*. 1. vyd. Praha : Národní knihovna České republiky, 2003. 43 s. ISBN 80-7050-415-3.
44. SYSEL, Jiří. *Koncepce managementu kvality : Systém managementu jakosti ISO 9000* [online]. 2006 [cit. 2007-03-10]. Dostupný z WWW: <<http://www.cestovni-ruch.cz/hotelieri/iso9000d.php>>.
45. Systém řízení kvality podle norem ISO 9000 [online]. 2005 [cit. 2007-03-10]. Dostupný z WWW: <<http://management.jakosti.cz/rizeni-kvality/>>.
46. TELNAROVÁ, Zdeňka. Úvod do databází : Distanční výuková opora. 1. vyd. Ostrava : Ostravská univerzita, 2002. 105 s. Učební texty Ostravské univerzity.
47. Thomson Gale. *Gale Directory of Online, Portable, and Internet Databases* [online]. 1.4.2005 [cit. 2005-04-06]. Dostupný z WWW: <<http://library.dialog.com/bluesheets/html/bl0230.html>>.
48. VLASÁK, Rudolf, BULÍČKOVÁ, Soňa. Základy projektování informačních systémů. 1. vyd. Praha : Karolinum, 2003. 144 s. ISBN 80-246-0727-1.
49. VOŘÍŠEK, Jiří. Informační technologie a systémová integrace. 1. vyd. Praha : Vysoká škola ekonomická v Praze, 1996. 198 s. ISBN 80-7079-895-5.
50. VOŘÍŠEK, Jiří. *Strategické řízení informačního systému a systémové organizace*. 1. vyd. Praha : Management Press, 1997. 323 s. ISBN 80-85943-40-9.
51. XIE, Nick, MEAD, Nancy. *SQUARE Project : Cost/Benefit Analysis Framework for Information Security Improvement Projects in Small Companies*. 1st edition. [s.l.] : [s.n.], 2004. 40 s. Networked Systems Survivability Program. Dostupný z WWW: <<http://www.sei.cmu.edu/pub/documents/04.reports/pdf/04tn045.pdf>>.
52. *Z39.50 implementation experience*. Edited by Paul Over, William E. Moen. 1st edition. Washington : U.S. Government Printing Office, 1995. 123 s. s. NIST special publication 500-/ computer systems technology.
53. *Z39.50 Implementors Group* [online]. 2000 , 2002 [cit. 2007-03-29]. Dostupný z WWW: <<http://www.stk.cz/ZIG/>>.
54. Z39.50 MAINTENANCE AGENCY. *The Z39.50 Document* [online]. 2004 [cit. 2007-03-29]. Dostupný z WWW: <<http://www.loc.gov/z3950/agency/document.html>>.
55. *Z39.50 profil JIB* [online]. 2004 [cit. 2007-03-29]. Dostupný z WWW: <<http://info.jib.cz/dokumenty/profiljib.pdf>>.

- 
56. Zákon o převodech peněžních prostředků, elektronických platebních prostředcích a platebních systémech. In *Sbírka zákonů : Česká republika*. [s.l.] : [s.n.], 2002. s. 3135-3145. Dostupný z WWW: <<http://www.mvcr.cz/sbirka/2002/sb055-02.pdf>>.
57. ŽABIČKA, Petr. OAI-PMH: Protokol pro metadatovou interoperabilitu. In *Automatizace knihovnických procesů : Liberec 2003*. 1. vyd. Praha : ČVUT, 2003. s. 43-48. Dostupný z WWW: <[knihovny.cvut.cz/akp2003/sbornik/05\\_zabicka.pdf](http://knihovny.cvut.cz/akp2003/sbornik/05_zabicka.pdf)>.