

MANUSCRIPTORIUM JAKO DISTRIBUOVANÁ DIGITÁLNÍ KNIHOVNA

Zdeněk Uhlíř

Od prosince 2007 probíhá projekt ENRICH (<http://enrich.manuscriptorium.com>), který končí v listopadu 2009. Cílem tohoto projektu je rozšířit počet řádných obsahových partnerů digitální knihovny Manuscriptorium (<http://www.manuscriptorium.com>) o instituce v řadě evropských zemí (Dánsko, Island, Itálie, Litva, Maďarsko, Německo, Polsko, Rakousko, Španělsko) a získat asociované partnery v dalších zemích (Bělorusko, Moldávie, Rumunsko, Rusko, Srbsko ad.). Manuscriptorium se tak už stalo mezinárodním, resp. nadnárodním zdrojem a tematickým agregátorem rukopisů a starých tisků pro Europeana (<http://www.europeana.eu>), tzn. evropskou digitální knihovnou rukopisů a starých tisků. Zvětšený rozsah Manuscriptoria a zároveň jeho prohloubený obsah však znamená, že jsme museli znovu promyslet a upravit základní principy, na nichž Manuscriptorium spočívá, aby byl možný jeho další rozvoj.

Až do zahájení projektu ENRICH koncem roku 2007 bylo Manuscriptorium v zásadě centralizovaným zdrojem, jakkoli už tehdy jsme jako jednu z jeho základních idejí přijímali také možnost distribuovaného uložení dat. To však bylo v praktickém ohledu naprostou výjimkou a mělo to spíše testovací nebo poloprovozní než opravdu rutinní charakter. Už tehdy, tj. někdy v letech 2006–2007, bylo zřejmé, že s distribuovaným uložením dat je nutno počítat nejenom teoreticky a takřkajíc platonicky, nýbrž že je nutno usilovat o jeho praktické uskutečnění na rutinní organizační úrovni. Národní knihovna České republiky usilovala o shromáždění členů konsorcia a o získání evropského grantového projektu ENRICH mj. právě z důvodu rozvoje Manuscriptoria ve směru k distribuovanému uložení dat.

S centralizovaností, která byla až do nedávné doby praktickou, byť nikoli koncepční vlastností Manuscriptoria, souvisí rovněž homogenita, a to zejména homogenita metadatových, ale i datových standardů. Vzhledem k tomu, že Manuscriptorium vzniklo na základě podprogramu VISK6 – Memoriae mundi series Bohemica, v němž bylo financování podmíněno dodržováním standardu MASTER, resp. MASTER+ (*msnkaip.dtd*), a vzhledem k tomu, že v rámci tohoto podprogramu Národní knihovna České republiky zaručovala trvalé, tj. dlouhodobé uložení dat, prakticky užití jiného standardu nepřicházelo v úvahu, ačkoli

heterogenita dat (míněno hlavně metadat) se jako možnost nikdy nevyklučovala. S přistoupením nových velkých partnerů do Manuscriptoria v rámci projektu ENRICH se však heterogenita stala skutečností, kterou bylo nutno prakticky a rutinně řešit, nikoli o ní jen teoreticky uvažovat.

Distribuovanost, resp. centralizovanost, a heterogenita, resp. homogenita, v sobě zahrnuje jak aspekty formální či technické, tak obsahové či věcné. Z hlediska technického jde o organizaci digitálního prostředí tak, aby různé zdroje byly interoperabilní, tj. byly způsobilé komunikace v rámci jedné technické infrastruktury, jakož i schopné poskytovat standardizovaná data pro konverze a další zpracování, a to vše tak, aby výsledkem byla agregace v dalším zdroji a integrace v jednom uživatelském rozhraní, tzn. v Manuscriptoriu. Jako netechnik se těmto formálním záležitostem nebudu věnovat, protože to za mne kvalifikovaněji učiní jiní. Z hlediska obsahového jde nejprve o shromáždění metadat a jejich prostřednictvím dat, tzn. pramenů ke studiu písemného kulturního dědictví v širokém smyslu. To je podmínka nutná, ale nepostačující. Dále jde o jejich převedení na společného jmenovatele, jak se říká, tzn. jejich zpracování takovým způsobem, aby bylo možné vyhledávání v jejich celku za využití stejných principů, na základě stejného dotazu atd. A konečně jde o to, aby byla vytvořena a získána heterogenní data (v tomto případě zejména metadata) ve věcném smyslu, tzn. variantní či alternativní metadata týkající se téhož objektu (historického dokumentu) připravená z rozmanitých hledisek zájmu, v pohledu různých speciálních disciplín atp. Hlubou budoucností je zatím čtvrtý stupeň integrace, který by zohledňoval uživatelský požadavek multilingvalismu nikoli ve smyslu různých jazykových lokalizací, nýbrž ve smyslu multilingválního vyhledávání; ačkoli v případě přirozených jazyků v podobě jejich dnešního stupně vývoje je tato otázka již v zásadě vyřešená (viz třeba Google), v případě historických dokumentů je nutno uvažovat o několika vývojových stupních jednoho a téhož jazyka, což zatím přesahuje naše možnosti. S tím částečně souvisí i sémantický web a využití ontologií (protože předpokládá sémantické způsoby vyhledávání), zároveň se však týká jiné oblasti (protože chce k cíli, který před sebe klade integrace zdrojů, dojít jinými prostředky).

Nejprve tedy jde o agregaci na nejjednodušší rovině, tzn. o shromáždění metadat a jejich prostřednictvím dat. To v podstatě znamená, že metadata jsou centralizována a fyzicky uložena v Manuscriptoriu, zatímco data (typicky digitální obrazy, tj. kopie originálních historických dokumentů, ale také plné texty, tj. edice primárních historických dokumentů) zůstávají fyzicky umístěna na serveru partnera a jsou dostupná prostřednictvím hyperlinků z metadat

k datům. Ačkoli koncový uživatel takřikajíc do tohoto způsobu agregace nevidí, a mohlo by se tedy zdát, že je pro něj zcela irelevantní, přece jen pro něj nezaobíratelný význam má.

Tento význam je na rovině řekněme politicko-organizační, protože se týká kulturních politik jednotlivých partnerských institucí: kulturní politiky institucí (v naprostém rozporu s běžným míněním) totiž ani zdaleka nejsou v souladu, ba jsou naopak velmi rozbíhavé a vyznačují se kontroverzností až konfliktností. V tomto případě jde o dvě věci, které mohou mít – a zpravidla mají – dopad na koncového uživatele. Jednak – tvrdě, ale nikoli cynicky řečeno – veřejné instituce v oblasti kultury potřebují svým zřizovatelům prokázat, že jejich činnost má smysl, že tedy zaslouží být financována. Proto se potřebují prezentovat, aby byly známy koncovým uživatelům. Integrace v jednom uživatelském rozhraní, neřku-li centralizace dat, jde tudíž proti jejich nejvlastnějším zájmům; kamenné instituce zkrátka dost dobře nesnášejí svěží vítr virtuálního prostředí. Jestliže však je agregace založena na distribuovaném uložení dat, potom lze snadno odkázat nejenom na domácí stránku příslušné instituce, která data vlastní či spravuje a k integraci je poskytla, ale také na stránku, kde sama tato data zpřístupňuje neintegrováná do širšího zdroje. Tímto manévrem se dostojí jak zájmům koncového uživatele, který preferuje data integrovaná, tak dílčího poskytovatele, která nechce zaniknout za jedním uživatelským rozhraním integrovaného síťového prostředí.

Jednak tu jde o vlastnictví dat. Digitální obrazové kopie originálních historických dokumentů jsou totiž chráněny autorskými právy, poněvadž nejde jen o jinou reprezentaci těchto dokumentů, ale v podstatě o jejich nové vydání. Jde tu tudíž o otázky distribuce těchto „staronových“ dokumentů. Vzhledem k tomu, že pro zpřístupnění na internetu jsou digitální obrazové kopie zpravidla poskytovány v komprimované podobě a nižší kvalitě a že obrazy v nejvyšší, archivní kvalitě zůstávají u poskytovatele a agregátor k nim nemá přístup, jakož i vzhledem k tomu, že reprodukce tiskem je obvykle možná jen z oněch obrazů nejvyšší kvality, nemusí se při distribuovaném zpřístupnění poskytovatel obávat zneužití svých dat, neboť je má plně pod kontrolou. Lze tak opravdu snadno zajistit, že agregátor zpřístupňuje data partnerů-poskytovatelů pouze na internetu, k čemuž jedinému je oprávněn, což je výhodná zýruka pro partnery-poskytovatele. Je to však i výhoda pro agregátora, protože se nemusí starat o bezpečné dlouhodobé uložení dat těchto partnerů.

Dalo by se tak říci, že distribuované zpřístupnění je výhodné jak pro agregátora, tak pro ostatní partnery. Nicméně je tu ještě další okolnost, na kterou

se obvykle nemyslí a která vypluje na povrch, až když nastanou problémy. Je to zvláštní a těžko pochopitelné, nicméně řada partnerů, resp. jejich vedoucích pracovníků si myslí, že integrace „se stane“ nějakým bližší neurčeným a kupodivu bezpracným způsobem. Až později nutně zjistí, že „se nestane“, protože v lidském světě se prostě nic neděje samo sebou. Ačkoli tedy vytváření distribuovaného integrovaného zdroje spočívá na malém počtu jednoduchých zásad, neobejde se to bez lidské práce, což znamená, že člověk/lidé určité kvalifikace/určitých kvalifikací se po jistý čas musí věnovat integraci. Pro technicky a obsahově vyspělé partnery je takové zatížení minimální, pro partnery ne tak vyspělé nebo třeba jen začínající je to příležitost jak prohloubit, nebo dokonce získat know-how. Agregace na bázi distribuovaného zpřístupnění a heterogenity standardů však obvykle znamená, že pro ni není zapotřebí přímých nákladů, tzn. skutečných, hotových peněz. V době krize státních rozpočtů ve většině zemí zabývajících se digitalizací je to na rozdíl od jiných poměrně povzbudivá skutečnost.

Druhou důležitou věcí při vytváření distribuovaného integrovaného zdroje je koordinace v oblasti popisných metadat. Pokud pro praktickou funkční agregaci dat jsou důležitá zejména metadata strukturální umožňující vytvářet vcelku libovolně komplexní digitální dokumenty v podobě virtuálních knih, pro věcnou či obsahovou sféru, v jejímž rámci se odehrává vyhledávání, tzn. kladení (a nesmíme zapomenout také na úspěšné zodpovídání) uživatelských dotazů, jsou závažná právě metadata popisná. Při různosti užívaných standardů mezi jednotlivými partnery Manuscriptoria je totiž nezbytná korelace mezi nimi, tzn. mapování, které strukturální jednotce (poli nebo elementu) jednoho standardu věcně odpovídá ta či ona strukturální jednotka (pole nebo element) aktuálního vnitřního standardu Manuscriptoria, tj. TEI P5 ENRICH (enrich.dtd). Na základě tak provedené korelace mezi jednotlivými standardy či formáty lze vytvořit vyhledávání a/nebo konverzi do vnitřního formátu Manuscriptoria. Pro úspěšnost vyhledávání (pro jeho lepší výsledky), jakož i pro jeho rychlost (vlastnost vyhledávacího stroje pro koncového uživatele nesmírně důležitá) se v Manuscriptoriu provádí konverze, takže v centralizované databázi metadat jsou všechny katalogové záznamy v jednom a téměř formátu. To pak dovolí indexovat všechna popisná metadata podle celé řady strukturálních jednotek (polí či elementů) a v důsledku toho spolehlivě a rychle vyhledávat.

Ale přitom je třeba uvědomit si jednu věc a nikdy na ni nezapomínat: vyhledávací stroj může vyhledat a koncovému uživateli poskytnout pouze a jen ty údaje, které jsou v popisných metadatach obsaženy. Kde nic není, ani smrt nebere, jak se říká – a informačně komunikační technologie jsou ke všemu mno-

hem slabší než smrt, i když na to někteří nepomýšlejí. A tak je nutno si znovu uvědomit, že záleží nejenom na kvalitě, tzn. především věcné spolehlivosti, ale také na informační hloubce, tj. intenzi popisných metadat. A zde samozřejmě tvorba jakéhokoli integrovaného zdroje není bez problémů, tudíž se to týká také Manuscriptoria. Tím se dostáváme k dalšímu aspektu heterogenity metadat: heterogenita, jak právě zjišťujeme, se netýká jen formální stránky (různé standardy či formáty), ale zasahuje také stránku obsahovou (různá informační hloubka, tzn. podrobnost). Zkrátka počítač se na jedné straně liší od člověka, že je jak mnohem přesnější než člověk, takřka bezchybný, tak mnohem rychlejší, avšak zároveň má s člověkem společné to, že jak jeden, tak druhý nenajdou nic tam, kde není co hledat.

Otázka obsahové heterogenity má tudíž různé aspekty, z nichž jsou nejdůležitější tři. Jako první je třeba uvést až alarmující informační ngramotnost koncových uživatelů. Ačkoli v případě Manuscriptoria jde zpravidla, ba takřka výhradně o vědecké a odborné pracovníky, tedy lidi vysoce kvalifikované, nelze si nepovšimnout, že jejich vysoká kvalifikace a hluboké chápání některých problémů jsou zhusta svým rozsahem natolik omezené, že takový koncový uživatel nakonec v všechno o ničem, jak se ironicky říká, že je totiž zjevným fachtidotem a v tomto ohledu se nijak neliší od mnohem méně vzdělaného tzv. člověka z ulice. Takovým koncovým uživatelům (a opravdu jich není právě málo) je naprosto nedostupné už obyčejné kombinované vyhledávání, neřku-li použití různých operátorů, takže na složitější expertní vyhledávání s možností vyloučení z předchozího výsledku nebo naopak přidání k němu je lépe se ani neptat. Ačkoli rozlišení mezi počítačovou a informační gramotností se na jedné straně stalo takřka folklorem, na druhé straně značná část koncových uživatelů Manuscriptoria o tom stále nic neví a je ve skutečnosti schopna vyhledávat jen na základě vepsání dotazu do jednoho jediného pole. Bizarní pak je, že právě těmto uživatelům Google používající (zajisté nejenom) jediné pole pro položku dotazu připadá primitivní a nedůstojný. Zde jsou tedy evidentní faktické meze, které zasahují mnohem širší oblast, než abychom s nimi v Manuscriptoriu mohli vůbec něco udělat, protože se to kupodivu netýká jen nejstarší generace, jak bychom si mohli myslet.

Jsou tu však i dva aspekty obsahové heterogenity, které se týkají přípravy dat, a s těmi něco udělat je poněkud snazší, třebaže nikoli automaticky. Nejprve je nutno přistoupit k pořizování metadat, ke katalogizaci z praktického hlediska. Je nutno si nejenom uvědomit, ale stále znovu to opakovat, že jakákoli data (platí to tedy i pro popisná metadata) existují jen natolik, nakolik jsou dostupná

v digitální podobě. Tudíž data dostupná jen v tištěné, strojopisné či rukopisné podobě pro digitální prostředí neexistují, poněvadž nejsou dostupná v digitální podobě – a v případě popisných metadat nadto platí, že v plné míře jsou v digitální podobě dostupná jen jakožto elektronický text, že tedy pouhý digitální obraz nestačí už jen z toho důvodu, že v něm nelze vyhledávat. Při ohromném množství dokumentů, které připadají v úvahu pro integraci do rozhraní Manuscriptoria a jež jsou do něho opravdu integrována, to pak znamená, že leckdy – tehdy, když informačně hlubší popisná metadata jsou dostupná jen v tištěné podobě, nebo když dokonce nejsou dostupná vůbec v žádné podobě – je nutno urychleně pořídit alespoň velice stručné a za jiných okolností naprosto nedostatečné záznamy. Méně je totiž někdy více: nejprve je nutno data s metadaty do digitálního prostředí vůbec dostat a teprve potom je možno uvažovat o informačním prohlubování popisných metadat. Jak se ukazuje, je tento praktický aspekt obsahové heterogenity záležitostí povýtce manažerskou, totiž otázkou organizace práce: ukazuje se, že tento problém je třeba řešit ve dvou etapách přípravy popisných metadat, tj. katalogizace, nejprve vytvářením inventárních záznamů a teprve po dokončení této etapy zahájit druhou etapu prohlubování záznamů (příčemž tu je někdy možno postupovat výběrově).

A tím se u obsahové heterogenity dostáváme na teoretickou rovinu. Je řada knihoven, resp. paměťových institucí, a jsou i takové mezi partnery Manuscriptoria, které nedisponují dostatečně kvalifikovanými silami, což se týká zejména katalogizace středověkých a raně novověkých rukopisů. V některých případech sice existují pomůcky, repertoria apod., které mohou tento nedostatek nahradit, ale to platí opravdu jen do jisté míry. Jednak jsou tyto pomůcky vzácné a drahé, a tudíž mnohdy k využití nedostupné, jednak jsou namnoze problematické tím, že jsou většinou postavené na konceptuálně ne zcela korektních základech. To je téma na hlubší a fundovanější diskusi (ostatně Richard Sharp o tom před časem napsal celou knihu), nicméně stručnou charakteristiku oné problematickosti lze poměrně jednoduše shrnout do jednoho dvojitého bodu. Doba „písemnictví bez autorů a bez generací“, jak středověk charakterizuje Antonín Škarka, a doba, která neznala pojem díla v dnešním významu, jak jsem se to pokusil doložit já, je ve většině takových repertorií prezentována v podobě soupisů děl podle jednotlivých autorů. Ale nedostí na této inkonsistenci; část děl uvedených v repertoriích (řada?, většina?), jak zjišťuje právě Richard Sharp, nenesou titul, který by byl doložitelný v pramenech, resp. není spojována s autorem, který by byl doložitelný v pramenech, tzn. tato díla nejsou tzv. evidence based. Tím nechci říci, že bychom neměli být vděční za dosavadní repertoria, protože je

to každopádně lepší, než kdybychom neměli nic. Je však radno si uvědomit, že taková repertoria prospějí spíše hotovému znalci příslušené problematiky než začátečníkovi nebo univerzalistovi písemného kulturního dědictví. Kámen úrazu v tom nebude jenom tehdy, když si uvědomíme dopad tohoto problému na budování zdrojů, a zejména integrovaných zdrojů, jako je Manuscriptorium.

Pokud jde o technickou stránku, je zcela zřejmé, že každý integrovaný zdroj musí být založen na kooperaci mezi partnery, přinejmenším mezi agregátorem/koordinátorem a jedním každým partnerem. Ukazuje se, že podobný způsob bude žádoucí i v případě obsahové stránky. Poněvadž se digitální knihovny nesnaží vytvářet nic takového, co by jen vzdáleně připomínalo catalogue raisonné, tzn. katalog určený pro samostatné čtení či samostatné čtení alespoň umožňující, není nutné ani předpokládat, že katalogizační záznam je jakousi mikromonografií rukopisu, atp. Z toho ovšem plyne, že není nutné požadovat, aby autorem jednoho záznamu byl právě jeden člověk. Připustíme-li však, že původci záznamu může být více lidí, potom je zde místo jak pro jakéhosi univerzalistu písemného kulturního dědictví, tak pro různé specialisty podle tematických, geografických, oborových nebo nakonec i kulturních okruhů, kteří se mohou podílet na pořizování jednoho jediného záznamu. Toto je zatím pouhá teoretická úvaha, protože dosavadní praxe je úplně jiná. Nicméně vícekolovost přípravy popisných metadat, kterou jsme identifikovali jako vhodný způsob práce při budování integrovaných zdrojů, nutně vede k úvahám v tomto směru. A to nemluvíme o rozvíjení technologií a metodik webu 2.0, které to po řádné autentizaci dovolují i osobám mimo přímý okruh paměťových institucí; jednoznačně k takovým postupům směřuje např. celá univerzitní a akademická sféra. I zde se samozřejmě musí nejdříve vybojovat kontroverze mezi kulturními politikami paměťových institucí (a nakonec dost možná i jednotlivců – proto nutnost řádné autentizace!), nicméně kooperativní přístup je na pořadu dne.

S tím souvisí třetí důležitá věc, totiž obsahová heterogenita popisných metadat ve vlastním smyslu, tedy nikoli už ve směru prosté informační hloubky, nýbrž z různých hledisek či pohledů. Tím se míní zejména katalogizační záznamy s ohledem na různé specializované disciplíny atp. V takovém případě by s největší pravděpodobností bylo nežádoucí hromadit různé mnohdy docela disparátní údaje ve fyzicky jediném záznamu. Takto vytvořený záznam by byl jednak informačně značně přetížen (a tudíž by po technické stránce byl docela obtížně zpracovatelný), jednak zatěžoval koncového uživatele, protože by mu předestíral celou řadu údajů, o které nejenom že nemá zájem, ale které mu navíc výrazně ztěžují orientaci. Nový systém Manuscriptoria, vzniklý částečně v sou-

vislosti s projektem ENRICH, dovoluje do jednoho komplexního digitálního dokumentu zařazovat více popisných záznamů, tzn. zařazovat různá popisná metadata ve variantní podobě podle jednotlivých speciálních zájmů a z různých speciálních hledisek. V tomto případě je pak naprosto evidentní, že různé varianty popisných metadat budou připraveny různými osobami, takže kooperativní způsob práce při budování integrovaného zdroje je v tomto případě nabíledni. Lze dokonce předpokládat, že prosazení takového postupu bude mnohem snazší než tvorba jediného katalogového záznamu více osobami.

Již od počátku práce na projektu ENRICH byl tým Manuscriptoria někdy kritizován za to, že orientuje Manuscriptorium příliš mezinárodně a nadnárodně a že více méně opomíjí úroveň národní. Tato kritika, ozyvající se především z Moravské zemské knihovny v Brně, má racionální základ v tom, že ne vždy všechno funguje tak, jak by mělo, a že zájem věnovaný až dosud domácím českým partnerům jsme museli více méně omezit ve prospěch partnerů zahraničních. Na druhou stranu však nelze nevidět, že tato kritika v sobě skrývá i myšlenky, které nejsou zcela v souladu s budováním integrovaných distribuovaných zdrojů. Považuji za nutné zdůraznit zde hlavně dvě věci. Přijatý princip heterogenity, který jediné dovoluje budovat integrovaný zdroj většího rozsahu, nedovoluje na druhou stranu vytvářet alespoň relativně homogenní prostředí tím, že by přes míru nutil partnery dodržovat jednu konkrétní podobu standardu, když standard sám připouští více různých podob. Manuscriptorium naopak nemá zájem na takové sjednocovací politice, protože je v širším měřítku prakticky neproveditelná. Ba navíc se zdá, že v jiné podobě tu ožívají kontroverze mezi kulturními politikami jednotlivých institucí, které jsou spíše překážkou spolupráce než jejím hybatelem.

Zdůrazňování důležitosti národní funkce Manuscriptoria však v sobě skrývá ještě jedno úskalí. O tom, zda se spíše chlubit kvalitními výsledky dosaženými na národní úrovni, anebo zda se prosazovat na úrovni mezinárodní či nadnárodní, je sice možno dlouho a zevrubně diskutovat, avšak pro další rozvoj to má pramalý význam, jak se domnívám. Je se však možno obávat, že zdůrazňování národní funkce Manuscriptoria a úlohy Národní knihovny České republiky v jejím naplňování znamená spíše vědomou či nevědomou snahu některých větších českých partnerů Manuscriptoria skrytě přenášet na Národní knihovnu České republiky ty činnosti, které ve skutečnosti mají vykonávat tito partneři. S tím poměrně jednoznačně souvisí představa, že se počítají pouze přímé náklady, protože tam jsou zapotřebí skutečné, hotové peníze, zatímco nepřímé náklady se nemusí počítat, protože tak skutečných, hotových peněz zapotřebí není. Je sice

pravda, že v případě nepřímých nákladů skutečných, hotových peněz zapotřebí není, nicméně to vůbec neznamená, že není zapotřebí činností, prostě práce, která se v těchto nepřímých nákladech vyjadřuje. A tato práce by podle představ některých českých partnerů do větší či menší míry měla padnout na Národní knihovnu České republiky, ačkoli podle toho, že se týká jejich dat, by ji měli dělat sami. Tady lze jenom říci, že při budování integrovaného zdroje je třeba být důsledný: ten partner (ať už domácí, nebo zahraniční), který se snaží být plně samostatný, se také musí připravit na to, že ponese důsledky své samostatnosti, tzn. přinejmenším v tom, že nebude přenášet své nepřímé náklady na agregátora a koordinátora. Budování integrovaných zdrojů prostě není procházka růžovým sadem, nýbrž práce, jejíž tíži musí každý zúčastněný chtít nechtě nést a nepokoušet se zatěžovat jiné partnery.

O Tvrdé věci, totiž sémantickém webu a ontologiích, mohu ztratit jenom pár slov, protože to je v nynější chvíli otázka budoucnosti. Představa ontologií je taková, že na rozdíl od tradičních selekčních jazyků, ať prekoordinovaných, postkoordinovaných či koordinovaných, neoperují s pouhými jednotlivými termíny či jejich kombinacemi, nýbrž s celými provázanými sítěmi těchto termínů. Dosud se ukázalo, že není možný jediný systém ontologií pro všechny oblasti lidského poznání, ale že musí existovat různé systémy pro dílčí výšece lidského poznání. I když si to někteří, mezi nimi zejména tradičně myslící knihovníci, nechtějí přiznat, přece jenom je jasné, že dokonalý jazyk, tedy to, co se snažily či snaží vytvořit všechny selekční jazyky, je principiálně nemožný, jak velmi stravitelně vyložil třeba Umberto Eco. Na druhou stranu je zřejmé, že v dílčí oblasti mohou mít ontologie opravdový heuristický význam. Potud poměrně jasná teorie. Potíže však nastávají při praktické implementaci ontologií, neboť se zásadně liší stanoviska, jak kýženého efektu dosáhnout. Někteří myslí, že ontologie jsou prostě slova, která se jako další metadata přiřadí k metadataům popisným, tzn. prostě se přiřadí k dokumentu, resp. k jeho plnému textu, aniž však jsou jeho součástí. V tomto případě však není zřejmé, čím podstatným se ontologie liší od tradičních selekčních jazyků. Jiní myslí (domnívám se, že správněji), že ontologie jsou slova, která se nacházejí v plných textech dokumentů, která z nich byla nějakým způsobem extrahována, vzájemně zřetězena, sestavena do souboru, resp. souborů, a tyto soubory jako jakési varianty implementovány do vyhledávacího stroje. Jakmile uživatel zadá vyhledávání některého z termínů ontologií uvedených v příslušném souboru, resp. příslušných souborech, vyhledávací stroj připojí další, která jsou s ním zřetězena, provede vyhledání na tomto komplexním základě a výsledkem bude – či spíše má být – sémantické

kontextuální vyhledávání. Potud teorie. V praxi jsou přinejmenším dvě úskalí, jednak obrovská masa práce při výběru a sestavování příslušného souboru, resp. příslušných souborů ontologií, jednak to, že musíme počítat s multilingválním vyhledávání nejenom v tradičním významu, tzn. mezi různými jazyky, ale také v poněkud netradičním významu, tj. mezi jednotlivými historickými stupni jednoho a téhož jazyka. Podle souhlasného mínění je tento problém dnes zatím neřešitelný. Představa sémantického webu, krátce řečeno, je potom taková, že bude existovat služba, resp. budou existovat služby, které budou na základě využití systému ontologií poskytovat chytřejší vyhledávání, než dnes poskytuje Google. Naděje lze zajisté mít lečjaké, nicméně měřit skoro všechno Googlem mi přijde poněkud idiosynkratické. Každopádně vidíme, že k implementaci ontologií do Manuscriptoria povede ještě dlouhá cesta a že ideu sémantického webu bude nutno ještě vytříbit, aby byla prakticky použitelná.

Myslím, že Manuscriptorium tím, že se stalo distribuovaným integrovaným zdrojem, učinilo v posledních dvou letech velký krok kupředu a že tím vyvolalo mnoho inspirativních otázek, na které se budeme muset všichni pokoušet odpovídat.