

Heterogenita jako základ virtuálního prostředí

Zdeněk Uhlíř

Národní knihovna ČR, Praha
Zdenek.Uhlik@nkp.cz

INFORUM 2005: 11. konference o profesionálních informačních zdrojích
Praha, 24. - 26.5. 2005

Virtuální prostředí – míním tím informační, komunikační a znalostní prostředí – není samo o sobě ničím konkrétním. Nedá se tedy říci, že by virtuální prostředí bylo v přímém protikladu k prostředím jiným, která jsme až dosud poznali, pokud tím míníme jejich technicko-technologickou povahu. Pokud bychom tedy chtěli klást rovnítko mezi virtuální prostředí a prostředí informačně komunikačních technologií, prostředí **elektronické** či **digitální**, klamali bychom sami sebe. Elektronické či digitální prostředí není ještě svou pouhou technologickou povahou prostředím virtuálním. Je ovšem správné, když řekneme, že elektronické či digitální prostředí je také prostředím virtuálním potenciálně: virtualita je potencialitou informačního, komunikačního a znalostního prostředí vytvářeného informačně komunikačními technologiemi, počítačovými i mobilními sítěmi. Řečeno jinak, virtualita je možnou vlastností takového síťového světa, který se zakládá nikoli na atomech, nýbrž na bitech, jak říká Nicholas Negroponte. Na druhou stranu však pro to, aby takové síťové prostředí bylo zároveň prostředím virtuálním, je třeba vytvořit určité předpoklady a k tomu povede ještě dlouhá cesta. V tomto příspěvku se pokusím naznačit, jak bychom měli uvažovat, abychom se po této cestě mohli úspěšně vydat.

Heterogenitu prostředí – v tom i virtuálního – lze potom chápat jako jeho nestejnorodost. Takové heterogenní prostředí je opakem prostředí homogenního, tzn. nepoužívá nutně ani povinně pro stejné cíle stejných technických modulů, technologických prostředků a metodických postupů. Konceptuální rámec heterogenního virtuálního prostředí je tedy založen na flexibilní metodologii, která nemusí nutně vycházet z představy přesné korelace mezi předmětem a metodou, byť jakási přiměřená korelace mezi nimi je zajiště nezbytná. Vzhledem k tomu, že v takto koncipovaném virtuálním prostředí je možno jeden předmět pojednávat různými metodami a *vice versa* že pro jednu metodu lze nalézt různé předměty, je povaha informační a znalostní práce v jeho rámci vyhraněně interdisciplinární, ba docela transdisciplinární. A poněvadž základní povaha informační a znalostní práce spočívá ve strukturované kombinaci zjišťování faktů, jejich vyhodnocení a následné kontextuální interpretaci, lze říci, že pro virtuální informační, komunikační a znalostní prostředí je v jeho heterogenitě principiální **symbolická analýza** ve více předmětných rovinách, nikoli pouhá adice či kumulace jenom na jedné předmětné rovině, jak velí pozitivistická a strukturalistická tradice. Virtualita a heterogenita prostředí tak je implicitně i komplexitou. Jde o to, abychom se pokusili tento implicitní, tj. zatím jen potenciální komplexní přístup postupně převádět do explicitní podoby.

Virtualita a heterogenita v takovémto vymezení přirozeně přináší určitá úskalí, protože hrozí rozpadem organizovaného a smysluplného vnímání světa jako kosmu v jeho vnímání neorganizované a v chaos beze smyslu. Oddělení předmětu a metody, na němž se toto vymezení zakládá, totiž vede k potenciální disparátnosti a následné nesrovnatelnosti a konec konců k obtížím ve sféře poznání a vůbec poznatelnosti. Heterogenní zdroje, prostředky a postupy je tudíž třeba **integrovat** aniž se však zároveň homogenizují: místo představy homogenity, stejnorodosti je třeba následovat představu sourodosti, která se konkretizuje ve vzájemné **konvertibilitě** dat a **interoperabilitě** nástrojů a celých systémů, které pracují s daty

a umožňují jejich přetváření v informace, transformace a znalosti. Naprosto zřejmým požadavkem kladeným na integraci heterogenního virtuálního prostředí tedy je nutnost oddělení věcných údajů od nástrojů, kterými jsou zpracovávány, tzn. oddělení dat od softwaru. To se na první pohled zdá být naprosto banální, ale při hlubším pohledu se ukazuje, že tomu tak vůbec není. Informační a komunikační technologie totiž přebírají některé úkony, které bez jejich použití jsou úkony mentálními, např. vyhledávání podle abecedy v kartotéce, výběr údajů podle určitého parametru v tištěném dokumentu atd. Vzájemná konvertibilita dat a interoperabilita nástrojů a celých systémů je tak pro integraci heterogenního virtuálního prostředí sice podmínka nutná, ale nepostačující.

Prostá technicko-technologická integrace musí mít svůj protějšek v **přípravě dat**. Data musí být oddělena od nástroje, který s nimi dále pracuje, jak už jsem se zmínil. Běžně se tomu říká oddělení dat od softwaru. To znamená, že je nutno zavést představu **validity** a **kompatibility** dat. Validita znamená, že nezáleží na nástroji, jímž jsou data pořízena, ale na výsledku, který musí splňovat určité čistě formální požadavky tak, aby vytvořená data byla použitelná jak pro další technické zpracování, tak pro reprezentaci koncovému uživateli. Kompatibilita znamená, že validní výstupy procesu přípravy dat musí být jednak zpracovatelné dalšími nástroji, jednak vzájemně převoditelné, pokud jde o výstupy téhož obsahového typu, a různým způsobem reprezentovatelné. To ve svém důsledku znamená nutnost používání veřejně přístupných (a v tom smyslu otevřených, protože uživateli ovlivnitelných), nikoli proprietárních **standardů**. To je další nutná podmínka pro heterogenní virtuální prostředí, ale ani tak to stále není postačující. Ještě je třeba zavést podmínku **neutrality** dat. Neutralita v tomto případě znamená, že musí jít o pouhá surová data, která nejsou žádným způsobem spojena s petrifikovanými operacemi jejich využívání, jako je např. abecední řazení, deskriptorové rejstříkování, tj. řazení porůznu se vyskytujících slov pod obligatorní, závazné termíny vyjadřující čistě pojmové denotáty apod. Řečeno jinak to znamená, že informační aparát nesmí být přímou součástí datového modelu a jeho struktury, ale musí být dostupný dynamicky, tzn. výsledek musí být vždy znovu vytvářen na základě opětovného dotazu, nesmí být do dat vložen předem. Teprve pak je zaručeno, že datové modely a informační a komunikační nástroje nemusí být homogenní a že přece lze využívat zároveň různých datových standardů a technologických nástrojů podle plurality hledisek, protože mohou fungovat v relativním kontextuálním rámci.

Integrace heterogenního prostředí se velmi výrazně projevuje v propojování a postupném splývání počítačových a telekomunikačních sítí, jakož i elektronických médií. To má rozhodně nejenom čistě technický, ale také **sociální** význam, třebaže dnes ještě nejsme schopni jej v plné míře posoudit, protože se nacházíme teprve na samém počátku tohoto procesu. V této integraci sítí původně výhradně určených pro koncová zařízení různého typu se ukazuje pravý význam Internetu jako nově nastupujícího typického média proměněného dominantního informačního, komunikačního a znalostního prostředí. Vzhledem k tomu, že mobilní telekomunikační sítě umožňují kromě hlasových také datové přenosy, a také vzhledem k tomu, že technologie digitální televize (ale ostatně už televize analogové) dovoluje kromě masového vysílání z jednoho místa na ostatní příjemcká místa také dialogové kombinované přenosy mezi větším počtem různých míst a že – zejména v případě televize kabelové – právě tak umožňuje datové přenosy, je nutno uvažovat o různých **funkcích** dílčích integrovaných sítí. Zdá se, že využívání takto integrovaných sítí by v budoucnu mohlo záviset nejenom na konkrétním místě člověka v rámci integrované sítě sítí (totiž podle toho, odkud je která ze sítí dostupná), ale také na potřebě **informační hloubky** v konkrétní chvíli a na konkrétním místě, kde se právě člověk nachází.

Potřeba informační hloubky totiž není dána jednou provždy podle toho, zda je člověk doménový expert nebo naopak laik. Ani není dána pouze tím, zda potřebné znalosti již má nebo je teprve získává. Potřeba informační hloubky závisí rovněž tak na kontextu činnosti

nejenom v interdisciplinárním a transdisciplinárním, ale také na **tematickém rámci**. V tom se projevuje paradigmatická změna, k níž přispívají informačně komunikační technologie a jejich integrace. Potřeba různé informační hloubky zatím není příliš rozvinutá, poněvadž protože nové paradigma informační a znalostní práce není dosud jasně ustáleno a neproniklo v reflektované podobě do širokého povědomí. Nejsou tedy dostatečně jasné typy uživatelských požadavků, a tudíž ani konkrétních požadavků na technickou způsobilost, aplikace a design příslušných koncových zařízení. Bude naopak trvat ještě relativně dlouhou dobu, než se požadavky na této obecné úrovni ustálí tak, aby bylo možno využívat všeho rozumného, co integrované heterogenní síťové prostředí nabízí. Je tedy nutné vycházet zdola a virtuální prostředí postupně budovat na **modulárním principu**, a to jak v případě aplikací, tak datových modelů.

Tak vzniká představa **sduženého dokumentu**. Dokument je v tradiční informační vědě definován ve souvislosti obsahu a nosiče jako záznam na nosiči. Při tom záznam tvoří obsahovou jednotu, třebaže může být hlouběji strukturován, a nosič je relativně samostatným fyzickým, lépe materiálním objektem. Dokument je podle této definice petrifikací informace nejenom v tom smyslu, že je externí paměť, ale také tak, že je dokončený (informace se v čase nemění) a uzavřený (informace se v čase nerozšiřuje či nerozmnoužuje). Takováto definice je pro síťové elektronické či dokonce virtuální prostředí zjevně nevhodná: v obecném pohledu zde nejsou jasně vydělené materiální objekty, na nichž jako na nosičích by byla informace petrifikována (třebaže mohou být), tzn. definice zakládající se na nutné souvislosti obsahu a nosiče nevystihuje podstatný rys existence obsahové jednotky ve virtuálním prostředí. Je potom otázka, jak definovat dokument bez nutné závislosti na nosiči, protože obecná definice obsahové jednotky je nepochybně potřebná. Pracovně lze říci, že dokument v tomto ne-materiálním pohledu se skládá z jednoho nebo více **souborů**, které mají samostatnou identifikaci jakožto digitální objekty a které jsou vytvořeny na základě téhož nebo i různých formálních technických standardů bez ohledu na fyzický nosič, na němž se nacházejí (pevný disk, magnetický pásek, kompaktní disk, DVD atp.).

Sdužený dokument skládající se z jednoho nebo více souborů pak může existovat ve dvojí základní podobě. Tradičnější z nich je forma **kumulovaná**. Je však obtížněji technicky proveditelná, pokud jsou jednotlivé soubory skládající se v dokument vytvořeny na základě různých formálních technických standardů (definicí typu dokumentu, textové, obrazové, zvukové soubory atp.). Kumulovaný sdužený dokument je v podstatě replikou tradiční tištěné knihy-monografie v elektronickém prostředí. Jeho význam spočívá spíše v přípravě podkladů pro tisk (v tom smyslu lze hovořit o elektronickém publikování) než v přímé reprezentaci ve virtuálním prostředí (a v tom případě pak lze o elektronickém publikování mluvit je stěží). Kumulovaná forma sduženého dokumentu je spíše jen **transpozicí** tradičních postupů tištěného informačního, komunikačního a znalostního prostředí do prostředí elektronického než něčím jiným, protože dokončenost a uzavřenost dokumentu zůstává v podstatě zachována. Avšak vzhledem ke konzervativnosti typického uživatele se právě tato forma v poslední době velmi rozvíjí: typický uživatel se sice začíná přiklánět k Internetu jako k výhodnému prostředku komunikace informace, ale jinak nechce příliš měnit své návyky a především se brání změnám myšlenkových paradigmat, k nimž síťové elektronické prostředí potenciálně vede.

Druhou formou sduženého dokumentu je forma **komplexní**. Z čistě technického hlediska je snáze realizovatelná, vyžaduje však masivní technickou infrastrukturu. Nezáleží při ní totiž na fyzickém místě uložení jednotlivých souborů, které dohromady vytvářejí sdužený dokument, takže pro využívání je třeba rychlé průchodnosti sítě, robustních vyhledávacích strojů a spolehlivých komunikačních protokolů. Komplexní sdužený dokument (oproti kumulované formě sduženého dokumentu) tak není replikou tradiční tištěné knihy-monografie v elektronickém prostředí, nýbrž novým fenoménem. Jeho význam

spočívá v přímé reprezentaci ve virtuálním prostředí (v tom smyslu lze hovořit o elektronickém publikování) než v přípravě podkladů pro tisk (a v tom případě pak lze o elektronickém publikování mluvit je stěží). Komplexní forma sdruženého dokumentu je opravdovou **transformací** tradičních postupů tištěného informačního, komunikačního a znalostního prostředí do prostředí elektronického než jejich pouhým převedením, protože dokončenost a uzavřenost dokumentu tu nemusí být jeho podstatným rysem a také jím ve skutečnosti není. Typický uživatel vzhledem ke své konzervativnosti jí tudíž není příliš nakloněn: podvědomě cítí podstatné změny, kterým se chce vyhnout, protože nechce měnit své návyky a nechce měnit zakořeněná myšlenková paradigmatata a jejich změnu má v tomto případě zcela jasně před očima. Je však zřejmé, že virtuální prostředí se svým podstatným rysem heterogenity se bude odvíjet především odtud.

Jestliže komplexní sdružený dokument ztrácí znak dokončenosti a uzavřenosti, vyplývá to především z toho, že je poměrně snadné a přitom z hlediska informačního obsahu velmi výhodné měnit stávající a dodávat nové dílčí soubory, z nichž se skládá. Této možnosti se proto zcela přirozeně využívá, což ovšem vede k tomu, že informační, komunikační a znalostní prostředí nabývá podstatného rysu jakési nestálosti, kterou nelze hodnotit ani výhradně negativně, ani výlučně pozitivně. Nejprve je nutné naučit se v této nestálosti prostředí orientovat. Tak vzniká představa **přechodného sdruženého dokumentu**, reagující na podstatnou, potenciální nedokončenost a otevřenost komplexního sdruženého dokumentu. Nejde ani tak o to, že by se při některém z příštích přístupů dokument na své adrese nenacházel, ale o to, že to nebude též dokument: bude na téže adrese, ale nebude shodný co do svého znění a v důsledku toho i informačního obsahu. Konkrétní znění dokumentu má či alespoň může mít jen více méně efemérní trvání, petrifikovaná informace je v externí paměti uložena pouze dočasně, nikoli trvale (třebaže jakékoli trvání má jistě jen relativní, nikoli absolutní význam). Lze sice předpokládat, že nová informace bude hlubší co do obsahu a širší co do rozsahu, přece však na to nelze spoléhat jako na nutnost. Je sice jasné, že ne každá informace zasluhuje být trvale uchována, nicméně problém je v tom, že každá informace se nachází v bohatě strukturovaném kontextu, který je nutno ustavičně zpětně ověřovat. Existence přechodného sdruženého dokumentu však hrozí zpětnou neověřitelností informace, protože při příštím přístupu na touž adresu může být jeho znění jiné, než je to, které má být ověřeno.

Ukazuje se tedy, že otázka přípravy dat velice úzce souvisí s otázkou **uchování dat**. Je totiž žádoucí uchovat data tak, aby se bral ohled na jejich **variabilitu** (proměnlivost obsahu vzhledem k času), **alternabilitu** (proměnlivost formálního zpracování vzhledem k času) a **flexibilitu** (proměnlivost účelů využití vzhledem k času). Variabilita dat se týká obsahových údajů a jejich změn, zatímco alternabilita se vztahuje prohloubení formálního zpracování, tzn. jejich fragmentace či segmentace a flexibilita odráží způsobilost nástrojů, které je zpracovávají. Ukazuje se tak, že otázka uchování dat má dva hlavní aspekty: **statický**, který se týká surových údajů, a **dynamický**, který se vztahuje k práci se surovými údaji. Ty je nutno pojednávat a řešit odděleně, protože to ostatně vyplývá z principu důsledného oddělení dat a nástrojů, které s nimi pracují. Pokud má být virtuální informační, komunikační a znalostní prostředí plně způsobilé k různým činnostem, které si v něm lze představit, a informačně plně spolehlivé, je třeba otázku uchování dat vyřešit právě s ohledem na **trvání v proměnlivosti**. Právě to je ta nejpodstatnější paradigmatická změna, která před námi nyní stojí.

Z hlediska variability, tzn. proměnlivosti surových dat jde o to, jak uchovat různé **verze** přechodného sdruženého dokumentu. Nejde totiž ani tak o jednotlivé dílčí údaje jako prvky dílčích souborů (třebaže i to v některých případech může hrát roli), ale zejména o množinu souborů, z nichž se skládá přechodný sdružený dokument a o změny jejího rozsahu. V případě spíše praktických a provozních činností je řešení poměrně snadné, protože lze

uložit příslušnou verzi, kterou lze chápat jako výsledek vyhledávání, lokálně. Vzhledem k tomu, že praktické a provozní činnosti zhusta mívají právní důsledky, je třeba řešit místo lokálního uložení tak, aby mělo právní relevanci, což se může případ od případu lišit, nemůže to tudíž být předmětem na tomto fóru. Totéž se týká polotovarů výzkumného, vědeckého, odborného a profesionálního provozu, kde ovšem zpravidla odpadá podmínka právní relevance. Z obecného hlediska uchování dat a informace to není problémem. Zbývá ovšem celá rozsáhlá oblast verzí přechodného komplexního sdruženého dokumentu, na něž se odkazuje v konečných výstupech, které ostatně také mohou být verzemi přechodných komplexních sdružených dokumentů. Lokální uložení zde jednak nemusí mít dostatečnou relevanci, jednak je ze sféry, v níž se konečné výstupy pohybují, nedostupné, a tudíž zde není prakticky možná zpětná ověřitelnost informace. To je ve srovnání s fungováním tradičního tištěného prostředí závažný nedostatek, který virtuální prostředí v nynější podobě výrazně diskvalifikuje.

Řešení však lze nalézt v představě **virtuálního dokumentu**. Virtuální dokument je uchovanou podobou příslušné verze přechodného komplexního sdruženého dokumentu, která byla v konečném výstupu použita a na kterou se odkazuje. Při tom je zřejmé, že takový dokument nemusí být výsledkem pouhého jednorázového automatického vyhledávání, ale může být také kombinací několika vyhledávání. Nutnou podmínkou však zároveň je, aby takový virtuální dokument byl běžně přístupný (i když to nutně nemusí znamenat veřejnou přístupnost v obvyklém smyslu). To předpokládá vybudování sofistikovaného informačního systému v rámci síťového virtuálního prostředí, v němž bude modul pro vytváření takových virtuálních dokumentů, který bude v jiných svých modulech zároveň zdrojem dat a jenž bude způsobilý k přenosu dat také z jiných zdrojů. Vzhledem k tomu, že paměťová kapacita dnes přestává být problémem, je to spíše jen otázkou organizačního zajištění a personálního vybavení pro administraci takového systému. Pro spolehlivou funkci virtuálního badatelského prostředí však je nezbytné, aby v jeho rámci pro jednotlivé tematické okruhy a speciální obory vznikly výseče – tzv. **gridy**, které by přes svou heterogenní povahu mohly být relativně jednotně organizovány. Zbývá jenom vyřešit otázku vzájemného vztahu licencovaných a nelicencovaných zdrojů v rámci každého jednotlivého gridu. To v zásadě není obtížný technický problém, zato je to však velký problém politický: ani badatelská sféra, ani sféra knihovníků, informačních pracovníků a archivářů nejeví příliš ochoty změnit zavedené paradigma dělby rolí, ačkoli právě této změny je nyní nanejvýše zapotřebí.

S každým jednotlivým gridem však souvisí i řešení otázek alternability a flexibility. Z předchozího řešení otázky variability v souvislosti s představou virtuálního dokumentu plyne, že se sice mohou měnit surové údaje ve svém obsahovém vyjádření, že by se však nemělo měnit jejich vyjádření formální. To v podstatě znamená, že kódování podle definice typu dokumentu apod. by mělo zůstat stále stejné, tzn. tzv. **ideologie markupu** by měla zůstat zachována. Jenom v tom případě má význam uchovávat virtuální dokumenty, které se rovnají verzím přechodných komplexních sdružených dokumentů. V případě alternability dat, tzn. změny jejich fragmentace či segmentace se však kódování podle definice typu dokumentu apod. mění, tzn. tzv. ideologie markupu je jiná. Zpravidla to znamená, že se využívá více bohatství elementů a atributů, jež definice typu dokumentu nabízí, aby se tak zaručily možnosti detailnějšího vyhledávání, sofistikovanějšího zpracování, členitější reprezentace atd. Lze si potom představit jako efektivní takovou cestu, kdy surové údaje v jejich podobě variabilních dat budou třeba uloženy v **základní databázi**, zatímco sofistikované údaje v jejich podobě alternabilních dat se budou nacházet v **prezentační databázi**. A lze si nakonec představit, že sofistikovaná alternabilní data budou ze surových variabilních dat pořizována několikerým způsobem pro různé účely, že se tedy nad jednou základní databází bude nacházet ještě několik prezentačních databází. V přítomnosti se tak sice zpravidla neděje, protože virtuální prostředí je teprve ve svých skromných počátcích, takže se ještě

nedosáhlo kritické míry jeho naplnění variabilními daty. Pro budoucnost však je třeba s takovou strategií nepochybně počítat. Obsahová heterogenita virtuálního prostředí se tím samozřejmě bude zvyšovat.

Flexibilita dat se týká jejich dalšího zpracování speciálními nástroji podle konkrétních specifických požadavků koncového uživatele. Spočívá ve snadné možnosti přenosu dat z jednoho systému do druhého, jenž má jiné vlastnosti a který dovoluje provádět další úkony potřebné pro jejich využití a pro přetvoření dat v informaci a následně transinformaci a znalost. V rámci takového dalšího systému bude zpravidla potřebná jejich další ještě sofistikovanější fragmentace či segmentace, tzn. bude zapotřebí dalšího speciálního kódování, tzn. tzv. ideologie markupu bude znovu jiná, odlišná jak od variabilních, tak od alternabilních dat. Pak ovšem bude náležité i uchování takových flexibilních dat také pro budoucí použití ve **speciální databázi**. Obsahová i formální heterogenita virtuálního prostředí tím bude dále vzrůstat. Důležité je, aby příslušná variabilní, alternabilní i flexibilní data byla součástí jednoho gridu propojeného více různými komunikačními protokoly. Jakkoli ani toto se zpravidla v přítomnosti neděje, postup od variabilních dat přes data alternabilní až k datům flexibilním je nepochybně perspektivní. Nelze však pominout, že to nebude jednoduché, protože bude nutno vyřešit řadu technických, organizačních a politických problémů; ty politické budou z nich asi nejobtížnější, jak bývá pravidlem.

Má-li být pro virtuální prostředí zaručeno trvání v proměnlivosti, je třeba na všech stupních rozlišovat mezi **primárními daty** na jedné a **sekundárními daty** na druhé straně. Primární data jsou taková data, která jsou původní vzhledem ke svému vstupu do příslušného systému a jako taková si zaslouží uchování v plném rozsahu. To znamená, že primární data v tomto smyslu nejsou data uloženými u toho, kdo je vytvořil, pokud je nevytvořil přímo pro některý ze systémů heterogenního virtuálního prostředí. Primární data jsou surovinou pro daný systém, nikoli polotovarem vytvořeným v rámci činnosti tohoto systému. Sekundární data jsou na druhé straně taková data, která vstupují do dílčích nadstavbových modulů systému a jako taková zasluhují uchování jen výběrově ve virtuálních dokumentech na něž se odkazuje z jiných sdružených dokumentů. Jinak jejich uchování v rámci některého ze systémů virtuálního prostředí není zapotřebí, což ovšem nevylučuje možnost jejich uchování lokálního jakožto polotovarů pro další využití individuálním koncovým uživatelem.

Virtuální dokumenty, tzn. verze přechodných komplexních sdružených dokumentů sestávají ze všech těchto typů dat, tj. jak variabilních, alternabilních a flexibilních na jedné, tak primárních a sekundárních na druhé straně. Vzhledem k tomu, že soubory dat se mohou sdružovat do dokumentů na základě lecjakých kombinací, je nutno ustavičně vyhodnocovat, nikoli pouze to, zda takové kombinace mají smysl (protože smysl nakonec mohou mít i věci neskutečné), ale zda jejich smysl je vhodný (vzhledem k ideové intenci), správný (vzhledem k metodickému postupu) a pravdivý (vzhledem ke korelaci s vnější skutečností, nikoli pouze se sférou vědomí). Jen pokud virtuální dokument splňuje nároky vhodnosti, správnosti a pravdivosti, lze jej považovat za **virtuální realitu** korelovanou v nějakém stupni ke skutečnosti vnějšího světa. Pokud však nároky vhodnosti, správnosti a pravdivosti nespĺňuje, je pouhým **faktoidem**, který je sice vnímatelný a pochopitelný, ale ve skutečnosti klamný, protože data kombinuje nevhodně, nesprávně a nepravdivě. Vytváření virtuální reality je novou fenomenální skutečností, s níž zatím máme pramalou zkušenost. Vyžaduje nové paradigma, protože stará pozitivistická a strukturalistická paradigma rozlišení mezi virtuální realitou a faktoidem nedovolují. Toto nové paradigma však zatím není ustáleno, a tak máme většinou s chápáním virtuální reality potíže. Jakkmile se však potřebné nové paradigma ustálí, rychlému budování virtuálního badatelského prostředí už nebude stát nic podstatného v cestě.

Práci ve virtuálním badatelském prostředí a s virtuálními dokumenty lze přiblížit na hypotetickém příkladu. (Hypotetickém z toho důvodu, že je sice teoreticky možný, v přítomnosti však prakticky nerealizovatelný, protože virtuální prostředí není dosud

v kritické míře naplněno variabilními, alternabilními a flexibilními daty a protože se zatím nedostává vhodných nástrojů ke zpracování těchto dat.) Funkční požadavky na bibliografické záznamy hierarchizují dokumenty do vrstev **dílo** (work) – **vyjádření** (expression) – **provedení** (manifestation) – **jednotka** (item), přičemž někdy se dílo a vyjádření spojují do jediné vrstvy (workexpression). Všeobecně se předpokládá, že hierarchicky sestupně (od díla k jednotce) se struktura rozvětňuje, tzn. že dílo má více vyjádření, každé vyjádření představuje více provedení a každé provedení zahrnuje více jednotek. Právě tak se všeobecně předpokládá, že hierarchicky vzestupně (od jednotky k dílu) se struktura spojuje, tzn. že množina jednotek se sdružuje pod jedním provedením, množina provedení se spojuje v jednom vyjádření a množina vyjádření tvoří jedno dílo. Je zřejmé, že tyto předpoklady vcelku věrně postihují fungování tištěného informačního, komunikačního a znalostního prostředí. Nazvěme tuto koncepci **objektivní**. Právě tak je ovšem zřejmé, že fungování jiných prostředí – rukopisného a pravděpodobně ani síťového elektronického – správně nepostihují. Tam je naopak třeba vycházet z koncepce **virtuální**.

Jestliže je tudíž naším cílem vytvořit virtuální badatelské prostředí pro práci s historickými fondy, je nutno funkční požadavky na bibliografické záznamy transformovat. Je to nutné z toho důvodu, poněvadž konceptuální rámec bibliografického záznamu nespočívá jen ve formalizaci dat, ale reflektuje také vztah dokumentů-textů-obsahových celků mezi sebou vzájemně. Středověké texty, tj. texty rukopisného období nevycházely pouze z (informačních) prameny, ale leckdy se přímo zakládaly na (datových) předlohách, které byly více méně doslova opisovány. Nejenom rukopisy jakožto materiální věci jsou v důsledku technologie své výroby individualitami, individualitami jsou také přepisované a opisované texty, protože každý písař v nich ve slabším případě prováděl redakční práci, v silnějším případě kombinoval předlohy a vytvářel nové znění textu, které zároveň bylo novým textem, třebaže to platí jenom relativně (absolutně však v dějinách neplatí vůbec nic). Texty zapisované do středověkých rukopisů tak měly více méně povahu **fluidního textu**. Použijeme-li hierarchizaci dokumentů dílo – vyjádření – provedení – jednotka, potom zjistíme že za takové situace se struktura rozvětňuje nejenom hierarchicky sestupně (od díla k jednotce, ale právě tak se rozvětňuje hierarchicky vzestupně (od jednotky k dílu). To znamená, že jedna jednotka může být subsumována pod více provedení a v důsledku toho také pod více vyjádření a nakonec děl. Pro tradiční pozitivistickou a strukturalistickou historii a filologii to je ovšem něco neslýchaného. Tzv. nová filologie to sice řeší kompromisně v tom smyslu, že každý rukopisný zápis je samostatným dílem, ale to – přestože je to přiměřenější – vystihuje věc pouze částečně. Je nutno přijmout metodologii fluidního textu, podle níž vyšší hierarchické stupně – vyjádření a dílo – jsou v zásadě virtuální.

Vzniká tudíž otázka, jak si s takovou situací jednoduše poradit tak, aby byla technicky snadno zpracovatelná a uživatelsky vhodně reprezentovatelná. Konceptuální rámec a datový model TEI zapsaný v XML umožňuje vhodně vyjádřit jak formální, tak obsahovou strukturu textu. Je tudíž možné zpracovávat pouze jeho části a kombinovat je s částmi jiných textů, a to jak v případě formální struktury (kapitoly, odstavce, věty apod.), tak také struktury obsahové v primárním (typická syntagmata, charakteristické slovní komplexy atp.) či sekundárním (citace, reference a aluze) ohledu. Na tomto základě je teoreticky velice snadné vytvářet virtuální dokumenty postihující virtuální podobu různých vyjádření a děl. V současnosti však to lze pouze mechanicky, „hlavoručně“, což neskýtá předpoklady pro širší využití v praktické badatelské práci, protože to je neúnosně časově náročné. V dalším výzkumu a vývoji je tedy třeba usilovat o možnost automatického zpracování, které by verze virtuálních komplexních sdružených dokumentů přímo generovalo a provedlo tak rutinní a hrubou práci. Vzhledem k tomu, že automatické generování vychází pouze z několika málo parametrů a nevyužívá jich komplexně, je ovšem vzhledem ke kontextu všech dokumentů více méně náhodné. Je tudíž následně nutno vyhodnotit, zda takto vzniklé verze virtuálních komplexních sdružených

dokumentů opravdu jsou virtuální realitou, anebo pouhým bezcenným faktoidem. To je ovšem samozřejmé, protože bez vyhodnocení zůstávají pouhá data, která nemohou být přeměněna v informaci.

Z celého předchozího výkladu plyne, že heterogenita virtuálního prostředí není jen záležitostí technickou a faktickou, ale že jde v rozhodující míře o záležitost konceptuální a paradigmatickou. V tomto směru bude nutno představu heterogenity dále rozvíjet.