

Evropský projekt ENRICH a jeho význam pro vybudování virtuálního badatelského prostředí

PhDr. Zdeněk Uhlíř / Národní knihovna ČR / zdenek.uhlir@nkp.cz

Resumé:

Autor zařazuje projekt ENRICH do širšího kontextu digitalizačních aktivit Národní knihovny České republiky, které jsou spojeny s digitální knihovnou Manuscriptorium. Přehledně popisuje všechny základní komponenty projektu i všechny jeho použitelné výstupy. Soustavně se věnuje i paradigmatickým změnám ve zpracování písemného a dokumentárního dědictví, které práce na projektu přinesla.

Klíčová slova: *písemné a dokumentární dědictví – digitální zpřístupnění – integrace zdrojů – virtuální badatelské prostředí.*

Summary:

Author puts the ENRICH project into a wider context of digitization activities of the National Library of the Czech Republic that relate to the Manuscriptorium Digital Library. He describes in a general survey all project work packages and all applicable outputs. He deals systematically with paradigmatical changes that work on the project brought for work with written and documentary heritage.

Keywords: *written and documentary heritage – digital access – resource integration – virtual research environment.*

Národní knihovna České republiky se účastní spolupráce na evropských projektech už od devadesátých let minulého století. Oblast historických fondů, resp. písemného a dokumentárního dědictví (*written and documentary heritage*) v tom nejenom neuzdávala stranou, ale zapojila se jako jedna z neaktivnějších. Zpočátku se pouze jako jeden z řadných partnerů projektového konsorcia účastnila práce na projektech MASTER (*Manuscript Access through Standard for Electronic Records* – <http://xml.coverpages.org/master.html>) v letech 1999-2001, VICODI (*Visual Contextualization of Digital Content* – <http://www.vicodi.org/>) v letech 2002-2004 a COMTOOCI (*COMputational TOOlS for the librarian and philological work in Cultural Institution* – <http://www.ilc.cnr.it/comtooci/>) v letech 2004-2005. Na základě získaných, relativně bohatých a pozitivních zkušeností se pak Národní knihovna České republiky rozhodla, že se nadále bude evropské spolupráce na projektech nejenom účastnit jako řadový partner, ale že se je pokusí i koordinovat. Došla totiž k poznání, že jenom tak může synergie partnerů různého typu z různých zemí plně využít. A tak postupně získala projekty ENRICH (*European Networking Resources and Information concerning Cultural Heritage* – <http://enrich.manuscriptorium.com/>) v letech 2007-2009, REDISCOVER (*Reunion of Dispersed Content: Virtual Evaluation and Reconstruction* – <http://rediscover.manuscriptorium.com/>) v letech 2009-2010 a EMBARK (*Enhance Manuscriptorium through BALcan Recovered Knowledge* – webová stránka v tuto chvíli dosud nebyla vytvořena) na roky 2010-2012. Z nich nejdůležitější je projekt ENRICH, ostatní dva jsou jenom jakýmsi doplňkem k němu.

Idea projektu ENRICH tedy nevznikla náhle, obrazně řečeno na zelené louce, nýbrž se zakládá na kontinuální systematické práci, kterou Národní knihovna České republiky digitalizaci písemného a dokumentárního dědictví od první poloviny devadesátých let minulého století věnovala. Jeho hlavním a nejdůležitějším cílem bylo integrovat digitální obsah, tzn. získat řadu nejvýznamnějších digitálních kolekcí písemného a dokumentárního dědictví z důležitých paměťových institucí většího počtu evropských zemí. Takové vymezení cíle projektu však nebylo dostatečné, protože bylo nutné vyřešit některé další problémy bezprostředně související s integrací dokumentů a sbírek, mezi nimi zejmé-

na zdokonalení standardů pro popisná, strukturální a technická metadata tak, aby byly obecněji použitelné, zavedení praktické metodiky integrace na technické a organizační rovině, vytvoření nástrojů pro poskytovatele primárních dat a personalizovaných nástrojů pro koncové uživatele, jakož i návržení prototypu multilingvální webové služby pro všechny kategorie uživatelů Manuscriptoria. Konečným cílem bylo vybudování základu pro evropskou digitální knihovnu rukopisů tak, aby byla způsobilá subagregovat data pro EUROPEANA a TEL (*The European Library*), respektive také pro The CERL Portal, případně další portály, metaindexy a další podobné integrované zdroje. Tento konečný cíl je zcela jasný, protože v dnešní době prosazujícího se síťového informačního a komunikačního prostředí bez jakékoli pochybnosti platí bonmot, že to, co není integrováno, neexistuje.

Práce na projektu ENRICH probíhala od 1. prosince 2007 do 30. listopadu 2009 a účastnilo se jí 18 partnerů z 12 zemí, a sice z Česka Národní knihovna České republiky (<http://www.nkp.cz/>) jako vědecký koordinátor a vedoucí komponenty 2 (tzv. *work package*), AiP Beroun, s. r. o. (<http://www.aipberoun.cz/>) jako technický koordinátor a vedoucí komponenty 5, Cross Czech, a. s. (<http://crossczech.cz/>) jako administrativní koordinátor a vedoucí komponenty 1, z Dánska Københavns universitet – Nordisk Forskningsinstitut (<http://nfi.ku.dk/>), z Francie Systran S.A. (<http://www.systran.fr/>) jako vedoucí komponenty 6, z Islandu Stornuj Árna Magnússonar í Íslenskum fræðum (<http://www.arnastofnun.is/>) a Landsbókasafn Íslands (<http://www.bok.hi.is/>), z Itálie Centro per la comunicazione e l'integrazione del media (<http://www.micc.unifi.it/>) jako vedoucí komponenty 4 a Biblioteca nazionale centrale di Firenze (<http://www.bncf.firenze.sbn.it/>), z Litvy Matematikos ir informatikos institutas (<http://www.mii.lt/>) jako vedoucí komponenty 7 a Vilniaus universiteto biblioteka (<http://www.mb.vu.lt/>), z Maďarska Budapešti Műszaki és Gazdaságtudományi Egyetem (<http://portal.bme.hu/>), z Německa Universität zu Köln – Historisch-Kulturwissenschaftliche Informationsverarbeitung (<http://www.hki.uni-koeln.de/>), z Polska Biblioteka inwersytecka we Wrocławiu (<http://www.bu.uni.wroc.pl/>) a Poznańskie Centrum Superkomputerowo Sieciowe (<http://www.man.poznan.pl/online/pl/>), z Rakouska Diözesanarchiv St. Pölten (<http://dasp.kirche.at/>), ze Španělska Biblioteca nacional de España (<http://www.bne.es/>) jako vedoucí komponenty 8, z Velké Británie Oxford University Computing Services (<http://www.oucs.ox.ac.uk/>) jako vedoucí komponenty 3.

Práce na projektu ENRICH sestávala z osmi komponent (*work packages*): 1. koordinace, management a administrace; 2. příprava pro implementaci systému a obohacení obsahu; 3. standardizace sdílených metadat; 4. personalizace pro koncové uživatele; 5. personalizace pro poskytovatele dat; 6. multilingvální a uživatelsky vlivný sofistikovaný přístup pro uživatele; 7. hodnocení, testování a ověřování; 8. šíření a využívání. První a poslední komponenta jsou obecné a jaksi obligatorní pro všechny evropské projekty v rámci kteréhokoli programu, ostatní (tj. druhá až sedmá) jsou specifické pro projekt ENRICH. Na základě spojení komponent různé úrovně bylo možno předpokládat, že výsledky a výstupy projektu ENRICH budou dobré, resp. alespoň uspokojivé, podaří-li se je vhodně skloubit. Bez nadsázky je možno říci, že se to podařilo.

Součástí první komponenty (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp1/>), kterou byly koordinace, management a administrace, bylo několik dílčích oblastí, a sice technologické řízení, projektová administrace a monitorování zdrojů, finanční koordinace, administrativní koordinace a konečně zajištění kvality. K tomu byly zřízeny výkonný výbor (*The Management Board*), technologická komise (*Technological Group Committee*) a výkonné komise jednotlivých komponent (*Work Package Management Committees*). Celkem se konalo sedm řídicích schůzek projektového konsorcia, a to 3. prosince 2007 v Praze, 28. února až 1. března 2008 v Kodani, 26.-27. června 2008 v Reykjavíku, 22.-23. září 2008 ve Florencii, 3.-5. prosince 2008 v Paříži, 26.-27. března 2009 ve Vilni a 9. června 2009 v Praze. V Madridu se ve dnech 5.-6. listopadu 2009 konala závěrečná konference, která byla věnována jednak propagaci projektu ENRICH, jednak prezentaci komplementárních projektů a programů, a to nejenom z EU, ale také z USA. Proběhly také národní workshopy; Národní knihovna uspořádala český workshop 19. listopadu 2009 v Praze a tým pracující v Národní knihovně České republiky na projektu ENRICH se zúčastnil také workshopů ve Vídni 28. května 2009 a

ve Vilně 26. března 2009. Proběhla i dvě hodnotitelská setkání v Lucemburku, průběžně 29. ledna 2009 a závěrečné 27. ledna 2010; v obou případech byla práce celého konsorcia na projektu ENRICH hodnocena velmi kladně. Díky úspěšné koordinační a organizační práci se podařilo ve všech zbývajících komponentách projektu splnit plánované úkoly, vytvořit funkční výstupy a dosáhnout výsledků, na kterých lze stavět i v budoucnosti.

Druhá komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp2/>), kterou byla příprava pro implementaci systému a obohacení obsahu, měla dílíč součástí jako vytvoření přehledu o dostupnosti digitálních kopií rukopisů a dalších historických materiálů, workshop technických partnerů týkající se postupu integrace, diskuse věnující se přípravě technických podmínek mezisystémové komunikace, jakož i definice standardů a vývoj metadat. Cílem komponenty bylo shromáždit veškeré dostupné informace o existujících digitálních zdrojích nejen všech řádných obsahových partnerů projektu ENRICH (*full partners*), ale i postupně se připojujících asociovaných partnerů, tj. těch, kteří vyjádřili ochotu spolupracovat na integraci dat i bez nároku na finanční podporu ze strany Evropské komise (*associated partners*). Získání těchto informací bylo důležité ze dvou důvodů. První z nich byl věcný, protože tak byl vytvořen podklad pro analýzu postupu agregace různých kolekcí jednotlivých partnerských institucí. Pečlivé shromáždění podrobných informací se ukázalo jako velmi prozíravé, protože z následující zevrubné analýzy vyplynulo, že reálně existující data nedovolují zvolit pro všechny partnery a kolekce jediný postup agregace, ale že je nutno podle povahy primárních dat postupovat čtyřmi různými cestami, kterými jsou: 1. automatický harvesting dat v případě technicky nejpokročilejších partnerů, 2. on-line konverze v případě partnerů průměrně technicky vyspělých, 3. off-line konverze v případě partnerů technicky méně pokročilých, 4. ruční příprava dat pro agregaci v případě partnerů s digitalizací teprve začínajících. Na základě takové detailní analýzy pak byla praktická agregace dat vcelku jednoduchou záležitostí. Druhý důvod byl organizační, protože údaje získané při předkládání žádosti o financování projektu se ukázaly jako nespolehlivé, byly totiž některými partnery velmi nadnesené z toho důvodu, že postihovaly veškerý digitalizovaný materiál, nejenom ten historický. Poněvadž však na základě těchto vstupních – a nespolehlivých – soupisů byly vtyčeny pro výstupy projektu ENRICH závazné kvantitativní výstupy, bylo nutné získat další data k agregaci odjinud. To se podařilo díky řadě asociovaných partnerů, mezi nimiž je nutno jmenovat zejména Univerzitu knižovnu Heidelberg v Německu (<http://www.ub.uni-heidelberg.de/>), Univerzitní knihovnu Lund ve Švédsku (<http://www.ub.lu.se/>), Národní knihovnu Rumunska v Bukurešti (<http://www.bibnat.ro/>), Národní knihovnu Běloruska v Minsku (<http://old.nlb.by/en/>), Trojisko-sergijevskou lavru v Rusku (<http://www.stsl.ru/manuscripts/index.php/>), jakož i ve Španělsku Knižovnu Univerzity Complutense (<http://www.ucm.es/BUCM/>) a Knižovnu královského paláce v Madridu (<http://www.patrimoniocultural.es/Home/Servicios-culturales-y-de-Investigacion/Real-Biblioteca.aspx/>). To sice znamenalo zvláště pro Národní knihovnu České republiky mnohem větší časové náklady než předpokládal plán projektu, avšak v konečném důsledku se to osvědčilo.

Třetí komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp3/>), která se věnovala standardizaci sdílených metadat, měla tři dílíč úkoly, a to: 1. vytvořit konverzi mezi platformami TEI P4 a TEI P5 pro popis rukopisů, 2. implementaci sklizení založeného na OAI-PMH do Manuscriptoria, a 3. rozšíření interního prostředí Manuscriptoria o kontejnerizaci METS v rámci schématu Manuscriptoria. Tato komponenta byla důležitá zejména v souvislosti se zdokonalením systému Manuscriptoria a s prohloubením jeho interoperability. Konverze mezi platformami TEI P4 a TEI P5 spočívala konkrétně v přechodu od staršího standardu MASTER vytvořeného v rámci stejnojmenného evropského projektu v letech 1999-2001, resp. od jeho extenze MASTER+ umožňující k popisným metadatům připojit i metadata technická a strukturální k TEI P5 ENRICH Specification, jež hlouběji integrovala rukopisný modul mezi ostatní moduly vytvořené v rámci TEI, přizpůsobila jej i popisu inkunábulí a starých tisků, resp. dalších tištěných dokumentů, a hlouběji integrovala strukturu komplexního digitálního dokumentu, přičemž připustila jeho vytvoření i v podobě distribuovaného komplexního digitálního dokumentu. K tomu byly vytvořeny i migrační nástroje vhodné pro okamžité nasazení v rámci projektové agregace. Velmi důležité je, že kromě vlastní TEI P5

ENRICH Specification byl vypracován i podrobný uživatelský manuál a další didaktické materiály s řadou praktických příkladů poskytujících návod pro konkrétní značkování v typických případech. Nemalý význam má i to, že v souvislosti se standardizací popisných a strukturálních metadat důležitých pro vytvoření uživatelsky vlídné virtuální knihy se v průběhu práce zároveň vynořila potřeba vytvořit i metodiku pro snadné vkládání nestandardních znaků do textových dat a tím usnadnit tvůrcům dat zpřístupňování historických textů v co nejautentičtější podobě bez zjednodušení, které ve většině případů vyžaduje zápis moderní grafikou s jejími sadami znaků. Šíře úkolů řešených ve třetí komponentě tedy v důsledku svého zaměření na praktickou užitečnost a použitelnost nakonec přesáhla plánovaný rozsah úkolů.

Čtvrtá komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp4/>) se soustředila na personalizaci pro koncové uživatele. Jejímí dílčími zaměřeními byla analýza typických potřeb koncových uživatelů, zejména badatelů a učitelů, s ohledem na implementaci typových kolekcí do Manuscriptoria, dále vytvoření individuálních virtuálních kolekcí pro koncové uživatele, vytvoření virtuálních dokumentů pro badatelské a pedagogické účely, vyhodnocení způsobilosti Manuscriptoria pro bibliografické vyhledávání, implementace možností hloubkového vyhledávání jak v metadatech, tak v textových datech a rozšíření schopností vyhledávacích strojů Manuscriptoria pracovat jak s metadaty, tak s plnými texty. Vzhledem k tomu, že každý digitální zdroj – ostatně také jako všechny zdroje tradiční – je určen především koncovým uživatelům, třebaže ani ostatní uživatelé nejsou přirozeně vyloučeni, byla tato komponenta velice významná, protože se soustředila na hlavní cíl, který má Manuscriptorium naplňovat, totiž sloužit především koncovým uživatelům, nikoli knihovnám a knihovníkům. U koncových uživatelů je třeba reálně počítat s velmi různou úrovní informační gramotnosti, a to – poněkud překvapivě – i v případě uživatelů vědecky a odborně vysoce kvalifikovaných. Proto kromě zdokonalení možností kombinovaného a pokročilého vyhledávání bylo nutno výrazně vylepšit i vyhledávání na základě jediné položky dotazu, poněvadž jenom takového způsobu vyhledávání je většina koncových uživatelů schopna. S ohledem na to nová platforma Manuscriptoria, založená na TEI P5 ENRICH Specification, nabízí koncovým uživatelům možnost využít takových položek dotazu, které se soustřeďují na vyhledávání specificky vymezených údajů za využití prvků přirozeného jazyka a zakládají se na strukturovaných logických konstrukcích, jež jsou uživateli neviditelné a nacházejí se uvnitř systému. Zároveň je u koncových uživatelů – zejména badatelů – nutno počítat s výrazně rozvinutými heuristickými potřebami a požadavky. Těm má v nové platformě Manuscriptoria odpovídat prostředí dovolující vytvářet statické (tj. na základě ručního výběru) a/nebo dynamické (tj. na základě uloženého dotazu) virtuální kolekce, které umožňují shromáždit a dále strukturovat pramenný materiál potřebný typicky k výzkumu či k výuce. Totéž platí pro virtuální dokumenty, které jsou typicky jakýmsi antologiemi jednotek různého typu (obrazů, stránek, intelektuálních jednotek, tj. textů) vybraných ze skutečně existujících digitalizovaných dokumentů. Prostředky pro práci s virtuálními kolekcemi a virtuálními dokumenty jsou zároveň prostředky, které vytvářejí základ pro elektronické publikování v digitálním síťovém prostoru. Práce, které byly v rámci čtvrté komponenty projektu ENRICH provedeny, se staly dobrým základem pro budoucí rozvoj Manuscriptoria jakožto důležitého zdroje všestranně zpřístupňujícího písemné a dokumentární dědictví, jakož i pro vytvoření virtuálního badatelského prostředí pro práci s písemným a dokumentárním dědictvím.

Pátá komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp5/>) se věnovala personalizaci pro poskytovatele dat. Dílčími úkoly bylo vytvoření on-line nástroje pro strukturování již existujících metadat a dat vztahujících se k rukopisům, analýza a vývoj nástrojů pro využívání rozsáhlých externích zdrojů dat, pilotní implementace rozsáhlých zdrojů dat poskytovaných vybranými partnery do Manuscriptoria a integrace externích dat do Manuscriptoria. V praktickém pohledu tedy pátá komponenta byla pro projekt ENRICH vůbec nejdůležitější, ba zásadní, protože z ní vzešel jeho hlavní výstup. Bez efektivní práce na komponentách druhé a třetí a bez výsledků této práce by tato komponenta vůbec nemohla být uskutečněna. Zejména se to týká shromáždění dat ode všech řádných i asociovaných partnerů, což ve skutečnosti znamenalo, že obsah Manuscriptoria se zvýšil zhruba pětikrát. Jelikož už před projektem

ENRICH bylo Manuscriptorium jednou z největších světových digitálních knihoven zpřístupňujících písemné a dokumentární dědictví, výsledky práce projektu ENRICH vedly k tomu, že se Manuscriptorium mezi obdobnými zdroji stalo rozhodně zdrojem největším, a to ve světovém měřítku. Shromáždění tak obrovské masy dat, navíc v podobě distribuovaných komplexních digitálních dokumentů, není vůbec triviální záležitostí. Vynutilo si přechod na novou, lepší a výkonnější platformu Manuscriptoria a řadu změn v jeho správním systému. Vynutilo si i flexibilnější přístup k datům a k jejich reprezentaci. Nebylo totiž, a ani následně v budoucnu není, možné vynucovat si homogenizaci standardů při přípravě a tvorbě dat u jednotlivých partnerů, protože všichni zpřístupňují svá data pro různé cíle a pro rozdílné účely, čemuž odpovídá používání různých standardů, jakož i jejich nestejná reprezentace. To se týká jak deskriptivních metadat, tak sekundárních dat (tj. dat zpřístupňovaných v Manuscriptoriu, nikoli dat uchovávaných v úložišti příslušného partnera, resp. v dílčím integrovaném zdroji, jež jsou data primárními), zejména variability různých úrovní kvality. Jako jediné možné se nakonec ukázalo redukovat úrovně kvality na pouhé tři, tj. galerii, náhledy a normální uživatelské obrazy. Nezbytným podkladem pro takovou flexibilitu jsou při praktickém procesu integrace konverze mezi standardy, jež se kromě formální technické stránky musí zakládat také na obsahovém mapování a korelování jednotlivých segmentů, elementů, atributů apod. Vzhledem k tomu, že všechny takové a podobné dílčí úkony byly v potřebném rozsahu náležitě provedeny, integrace všech zdrojů v rámci projektu ENRICH proběhla úspěšně bez problémů a výraznějších potíží.

Tématem šesté komponenty (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp6/>) byl multilingvální a uživatelsky vřidný sofistikovaný přístup pro uživatele. Součástí byly aktivity týkající se jednak vývoje multilingválního zpřístupnění, jednak vytvoření a praktického využívání překladového formuláře, dále implementace ontologií zakládajících se na výsledcích projektu VICODI, jehož se Národní knihovna České republiky v letech 2002-2004 účastnila jako jeden z řádných partnerů. V rámci komponenty multilingválního zpřístupnění tedy bylo nejprve do vyhledávacího rozhraní Manuscriptoria integrováno rozhraní pro programování aplikací (API – *Application Programming Interface*) zprostředkující multilingvální vyhledávání. Poněvadž systém Manuscriptoria je sám o sobě už tak dost složitý, byla zvolena varianta nezávislého multilingválního vyhledávání a nikoli stejně relevantní možnost multilingválního vyhledávání asociovaného se systémem Manuscriptoria. Tím bylo vytvořeno pro uživatele (včetně koncových uživatelů) překladové rozhraní umožňující automaticky prováděné překlady mezi řadou jazyků (angličtina, čeština, dánština, francouzština, islandština, italština, litevština, maďarština, němčina, polština, španělština) kontrolovat, korigovat a vylepšovat, jakož i provedené korekce a podané vylepšení zařadit do slovníků, resp. slovníků, na jejichž základě se vytvářejí automatické překlady. S integrací rozhraní pro automatické překlady a multilingvální vyhledávání do uživatelského rozhraní Manuscriptoria souvisí také vytvoření (resp. adaptace) překladového formuláře (STS – *Systran Translation Style-sheet*), založeného na aplikaci XSL transformací na původní XML dokumenty obsažené v databázi Manuscriptoria. Zároveň s překladovým formulářem byly využity i instance VICODI ontologií, ovšem jen nejelementárnějším negativním způsobem, kdy byly použity pro vyloučení překladů vlastních jmen (což se neobešlo bez obtížně řešitelných, resp. v danou chvíli vůbec neřešitelných problémů v tom případě, kdy vlastními jmény jsou slova, která jsou zároveň jmény obecnými). Celkově lze říci, že šestá komponenta projektu ENRICH měla spíše prospektivní a testovací než praktický a implementační význam. Ukázala spíše jen skutečný dosah problémů spojených s multilingválním překladem a multilingválním vyhledáváním, než aby dotyčné problémy efektivně vyřešila. Výstup této komponenty má spíše jen význam ukázkový, než aby byl opravdu důležitý pro koncového uživatele.

Sedmá komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp7/>) se zaměřila na hodnocení, testování a ověřování. Sestávala ze dvou dílčích úkolů, a to z vypracování evaluační strategie a dále z průběžného hodnocení dostupnosti, použitelnosti a přizpůsobitelnosti jednotlivých nástrojů či aplikací vyvinutých v projektu ENRICH. Vstupní část práce spočívala ve studiu základních principů a evaluačních kritérií vypracovaných široce známými řešitelskými týmy a v jejich adaptaci na hodnocení,

testování a ověřování nástrojů a aplikací, jež bylo plánováno vyvinout v rámci projektu ENRICH. V tomto přípravném stadiu evaluačních prací byly vytvářeny kategorie indikátorů podle stupně důležitosti a v jejich rámci jednotlivá dílčí kritéria a jejich váhy, rámec pro popis pozitiv a negativ hodnocených nástrojů a aplikací tak, aby mohl sloužit pro zdokonalení pracovních postupů a technik, jakož i rámec pro hodnocení platformy nově vytvářených dat (zejména automatického překladu) a jejich užitečnost pro nově implementované nebo modifikované nástroje. Ve vlastním evaluačním stadiu sedmé komponenty bylo primární soustředit se na technické aspekty a použitelnost systému Manuscriptoria, tj. jeho nové platformy, a to jakožto celku i s ohledem na jednotlivé jak adaptované a zdokonalené, tak nově vytvořené aplikace a nástroje. Hodnocení se týkalo všech aspektů komplexní služby Manuscriptoria a z obsahového hlediska všech jazykových oblastí. Výsledky hodnocení byly použity jako zpětná vazba pro technické partnery a jako podklad pro závěrečné ladění všech vyvinutých aplikací a nástrojů. Důležité bylo hodnotit nejenom interní složky systému Manuscriptoria, ale celý systém s ohledem na jeho interaktivitu a interoperabilitu, poněvadž konečným cílem projektu bylo sdílet informace napříč databázemi a jinými síťovými entitami. A v neposlední řadě bylo nutné testovat, zda obdobné datové modely a sady metadatových elementů jsou vskutku používány pro sémanticky obdobné jednotky a koncepty. Poněvadž žádný z partnerů podílejících se na práci na sedmé komponentě projektu ENRICH nepřistupoval k jejímu řešení formálně, zpětná vazba poskytnutá technickým partnerům byla velice účinná a v důsledku toho i aplikace a nástroje podléhající hodnocení jsou pro všechny typy uživatelů vhodné a snadno použitelné.

Osmá komponenta (<http://enrich.manuscriptorium.com/index.php?q=partner-area-wp8/>) se soustředila na šíření a využívání výsledků a výstupů projektu ENRICH, jakož i na získávání asociovaných partnerů a také partnerů, kteří přistoupí až po uzavření projektu ENRICH. Zahrnovala pět dílčích úkolů, jimiž byly: vytvoření plánu šíření, zhotovení propagačních materiálů, zřízení a udržování webové stránky, sestavení plánu využívání a konečné zorganizování závěrečné konference v Madridu v posledním měsíci řešení projektu ENRICH. Důležité bylo zejména zřídit na webové stránce projektu ENRICH sekci »How to Join Us«, ve které byly umístěny praktické informace pro nově přistupující partnery, aby spolupráce s nimi hned od počátku byla efektivní a přinášela konkrétní výsledky. Vzhledem k tomu, že při práci na rozsáhlých projektech a při budování agregovaných zdrojů, které mají mnoho partnerů, tj. poskytovatelů dat, je nutno pečlivě dodržovat formality, na stránce byl v první řadě umístěn vzorový dopis vyjadřující zájem účastnit se jako asociovaný partner projektu ENRICH a stát se obsahovým partnerem Manuscriptoria (*Letter of Intent*). Dále byl na webové stránce projektu v této sekci dostupný dotazník, na jehož základě je možno získat zevrubné informace o obsahu a formě kolekcí a dat, o užívaných standardech, o technických procesech souvisejících s jejich zpracováním, ukládáním a uchováváním, jakož i o otázkách týkajících se práv k datům; právě práva k datům mají z organizačního hlediska podstatnou důležitost, protože bez jejich náležitého vyřešení dost dobře není možné data ani agregovat, ani zpřístupňovat. A konečně jsou zájemcům o partnerství dostupné vzorové verze licenční smlouvy o poskytování dat pro Manuscriptorium, dodatku k licenční smlouvě, jakož i podlicenční smlouvy pro využívání Manuscriptoria. Toto praktické zaměření webové stránky projektu ENRICH se ukázalo jako velice užitečné a účinné, protože se podařilo získat ke spolupráci řadu dalších partnerů, a to i takových, kteří dříve váhali nebo jakoukoli spolupráci dokonce odmítali.

V důsledku toho bylo rozhodujícím výsledkem projektu ENRICH výrazné rozšíření počtu partnerů Manuscriptoria a takřka pětinasobné zvýšení množství integrovaných dat, pokud jde o počet dílčích analytických jednotek, tj. digitálních obrazů, resp. desetinásobné zvýšení, pokud jde o počet souhrnných jednotek, tj. komplexních digitálních dokumentů. Partneri, kteří aktuálně do Manuscriptoria dodávají komplexní digitální dokumenty, nejsou tedy jen z Česka, ale také z řady jiných zemí, konkrétně z Běloruska, Dánska, Islandu Itálie, Kazachstánu, Litvy, Maďarska, Moldávie, Německa, Polska, Rakouska, Rumunska, Ruska, Slovenska, Španělska, Švédska. Kromě toho někteří partneři z Chorvatska, Turecka a Ukrajiny dodávají pouze katalogové záznamy, resp. popisná metadata. A v současné chvíli probíhají více méně nadějná jednání s další-

mi partnery z Gruzie, Chorvatska, Slovinska a Srbska o připojení jejich komplexních digitálních dokumentů v distribuované podobě. Důležité pro rozšiřování sítě partnerů Manuscriptoria bylo také pořádání národních workshopů ve všech partnerských zemích – zejména španělských, polských, maďarských a rakouských partnerů byli ve svém úsilí iniciativní, takže se bezprostředně podařilo získat další partnery ve Španělsku a v Polsku a Národní knihovna České republiky na základě toho mohla efektivním jednáním získat další partnery v Maďarsku a pokusit se zahájit jednání v Rakousku, byť zatím bezvýsledně. Úspěšné šíření projektu ENRICH, jeho výsledků a výstupů ve všech účastnických zemích vedlo také k tomu, že Národní knihovna České republiky mohla v rámci nových projektů REDISCOVER (<http://rediscover.manuscriptorium.com/>) od září 2009 do listopadu 2010 a EMBARK (webová stránka zatím není zřízena) od září 2010 do dubna 2012 získat v prvním případě další partnery v Litvě, Polsku a Rumunsku a v druhém případě v Bulharsku, Řecku a Srbsku, přičemž ve většině případů je pravděpodobné, že pro Manuscriptorium poskytnou distribuované komplexní digitální dokumenty reprezentující písemné a dokumentární dědictví příslušných zemí. Aktivní přístup většiny řádných partnerů projektu ENRICH k plnění úkolů jeho osmí komponenty vedl k tomu, že byl položen základ pro budoucí perspektivu Manuscriptoria a pro další rozvoj sítě partnerů Manuscriptoria jak v nejbližší, tak i ve vzdálenější budoucnosti.

Synergicky s tím působí také ta skutečnost, že všechny výstupy projektu ENRICH byly směřovány do oblasti aplikovaného výzkumu, takže mají podobu standardů, technických nástrojů, technologických procesů a metodických postupů. Prvním z těchto výstupů je aplikace TEI a jejího rukopisného modulu v podobě TEI P5 ENRICH Specification (<http://tei.oucs.ox.ac.uk/ENRICH/>). Ačkoli je to primárně aplikace rukopisného modulu TEI umožňující spojení zejména popisných a strukturálních metadat do jediného komplexního dokumentu, dovoluje vytvářet také deskriptivní metadata vztahující se k inkunábulím a starým tiskům, resp. dalším starším tištěným dokumentům. To je nemírně důležité, protože použití TEI P5 ENRICH Specification tak jednak dovoluje překonat jistou archaičnost a antikvovanost MARCOvých formátů typicky a dominantně využívaných pro katalogizaci inkunábulí a starých tisků, která spočívá v tom, že MARCOvé formáty a knihovní systémy, jež s nimi pracují, jsou jen transpozicí tradičních lístkových katalogů do elektronické podoby a ničím více, jednak spojit všechny typy písemného a dokumentárního dědictví do jediného zdroje bez nutnosti předchozích konverzí, jež jsou nutně vždycky ztrátové. Ačkoli to může vypadat jako pouhá technická či technologická věc, ve skutečnosti se tím prosazuje paradigmatická změna, která je – ostatně jako každá paradigmatická změna vůbec – nezbytně kontroverzní. Místo důrazu, který tradiční kodikologie a bibliologie kladly na fyzický nosič informace, protože jej považovaly za rozhodující podklad pro evaluaci informace obsažené v tom kterém dokumentu, po této paradigmatické změně se za rozhodující podklad pro evaluaci považuje obsažená informace sama, tzn. hlavním požadavkem je *content, not container*. To má dopad na dosavadní tradiční knihovněd a všechny pomocné vědy historické, protože se tak důraz přesouvá ze studia vnějších znaků či formálních rysů na vnitřní znaky či obsahové rysy. Řečeno jinak, jako důležitější se začíná jevit analytický rozpis obsahu, nikoli detailní popis dokumentu ve fyzickém pohledu. A jelikož informace jakožto obsah dokumentu se pohybuje plynule na škále mezi konkrétním a abstraktním, je tedy svou podstatnou povahou virtuální, znamená tato paradigmatická změna zároveň přechod od zkoumání jednotlivého jevu ke zkoumání jevu hromadného. A poněvadž vědecké fakty v oblasti disciplín zabývajících se písemným a dokumentárním dědictvím jsou vlastně konstrukty a modely, nejsou tedy skutečností samou, potom každá – tudíž i tato – paradigmatická změna znamená, že to, co dosud bylo považováno za fakty, přestává jimi být, a že je naopak třeba nových, tzn. jiných fakt, jež jsou však v tuto chvíli sotva známa jen v konceptuální podobě, nikoli však konkrétně. *Summa summarum*, TEI P5 ENRICH Specification je základem pro podstatné změny, kterých dnes nejsme schopni dohlédnout.

Druhým výstupem je personalizované prostředí Manuscriptoria (<http://beta.manuscriptorium.com/apps/main/en/index.php/>), které kromě obvyklého využívání samotné digitální knihovny dovoluje nastavit si osobní účet a v jeho rámci vytvářet statické a dynamické virtuální kolekce. Možnost metodické práce s virtuálními kolekcemi je opět projevem paradigmatické změny, kterou využití informačních a komunikačních technologií

pro práci s písemným a dokumentárním dědictvím přináší. Statická virtuální kolekce spočívá v ručním výběru jednotlivých dokumentů z Manuscriptoria, resp. i z jiných zdrojů a v jejím umístění do vlastního personalizovaného prostoru, přičemž může být následně využívána pouze svým tvůrcem, anebo může být sdílána, a to buď přesně vymezeným dalšímu uživateli, anebo zcela všeobecně. Statická virtuální kolekce tedy má charakter dovolující jak uzavřenost, tak otevřenost, a to podle arbitrárního rozhodnutí jejího tvůrce. Má také charakter dovolující jak dokončenost, tak nedokončenost, a to právě tak podle arbitrárního rozhodnutí jejího tvůrce. Je však zřejmé, že inklinuje spíše k dokončenosti, protože její další čím dál tím organizačně náročnější doplňování bude ve stále větším rozporu se snižující se informační hodnotou přírůstků či doplňků. Dynamická virtuální kolekce naproti tomu spočívá v uložení dotazu a vždy novém vyhledávání dokumentů, které podle příslušného dotazu do ní patří, při každém jejím dalším vyvolání či otevření. Právě tak jako pro statickou virtuální kolekci pro ni platí, že může být následně využívána pouze svým tvůrcem, anebo může být sdílána, a to buď přesně vymezeným dalšímu uživateli, anebo zcela všeobecně. Dynamická virtuální kolekce má tudíž charakter uzavřenosti či otevřenosti podle vůle jejího tvůrce, čímž se shoduje se statickou virtuální kolekcí, jednoznačně má však charakter nedokončenosti, čímž se od statické virtuální kolekce odlišuje. A tak vidíme, že i virtuální kolekce jsou součástí paradigmatických změn, ke kterým dnes při zpracování a využívání písemného a dokumentárního dědictví dochází. Kromě efektivnějšího zpracování a využití vnitřních znaků či obsahových rysů a kromě snadného zkoumání hromadného jevu, který prostředek virtuálních kolekcí umožňuje, je tu ještě další charakteristický rys stávající paradigmatické změny, a to tendence k otevřenosti a nedokončenosti, v tomto případě zdroje. To má závažné důsledky pro konceptualizaci jak heuristických korpusů a tezaurů, tak pro konstruování faktů a modelů. A nadto vidíme, že práce s virtuálními kolekcemi přináší novou kvalitu do oblasti vědecké komunikace a elektronického publikování, ačkoli v tomto případě bude ještě nutno leccos dalšího udělat.

Třetím výstupem projektu ENRICH, který má velký význam pro budoucí práci s písemným a dokumentárním dědictvím, je MTool (<http://beta.manuscriptorium.com/apps/m-tool/m-tool.php/>), což je nástroj pro tvorbu distribuovaných komplexních digitálních dokumentů, jakož i virtuálních dokumentů, do něž je přímo implementována TEI P5 ENRICH Specification. To v sobě má přirozeně zároveň jak výhodu, tak nevýhodu. Výhodou je to, že MTool je tabulkový editor, jeho uživatel tudíž nemusí vytvářet značkování XML, ani je nemusí vidět, když nechce. To může mít a má svůj význam pro méně pokročilé a méně zkušené uživatele, protože TEI P5 ENRICH Specification je standard poměrně složitý, takže tvorba značkování není zcela triviální záležitostí. Výhodou je i to, že MTool zároveň vylišuje formální správnost vytvořeného XML dokumentu, jakož i náležitou dostupnost všech obrazů v rámci distribuovaného komplexního digitálního dokumentu. Jeho nevýhodou však je to, že právě z toho důvodu, že to je tabulkový editor, nevyužívá a ani nemůže využívat celé bohatství standardu při vytváření značkování, což znamená, že v některých případech nelze podat informaci tak, aby byla vyjádřena v datech segmentovaných do hloubky. Protože se však příprava dat používající do hloubky segmentované informace týká v podstatě jen pokročilých a zkušených uživatelů, není tato vlastnost MToolu neřešitelným problémem. V MToolu lze totiž sestavit jen základní schéma distribuovaného komplexního digitálního dokumentu s vložením cest k obrazům ve vzdáleném úložišti a popisná metadata vytvořit v některém ze standardních XML editorů a posléze je do schématu vytvořeného MToolem vložit. Další výhodou MToolu je to, že je to volně dostupný on-line nástroj, avšak jím vytvořené dokumenty se ukládají off-line na tvůrcův počítač. Nejsou tedy primárně součástí databáze Manuscriptoria, ale lze jich využít i jinde a jinak. I to je jistým způsobem paradigmatické změny oproti tradičnímu prostředí a tradičnímu způsobu práce, že totiž jednou vytvořená data lze libovolně mnohonásobně použít.

Čtvrtým výstupem je MCan (<http://www.dbase.cz/manuscriptorium/apps/candidates/login.php?envLang=eng/>), což je prostředí na testování MToolem vytvořených distribuovaných komplexních digitálních dokumentů. Je určeno zejména partnerům, kteří s digitalizací písemného a dokumentárního dědictví teprve začínají, anebo partnerům, kteří nemají popisná metadata, takže zpřístupnění jejich dat je problematické nebo vůbec

nemožné, či konečně partnerům, kteří jsou sice v digitalizačních aktivitách již zkušení, ale mají malé množství dat a užívají jiných standardů, takže nejspokladnější cestou pro ně, tak jako tak, je ruční tvorba distribuovaných komplexních digitálních dokumentů, poněvadž vytvoření konvertoru pro automatickou konverzi by v takovém případě bylo naprosto neekonomické. MCan není součástí personalizovaného prostředí pro koncové uživatele, je výhradně určeno již činným nebo alespoň potenciálním obsahovým partnerům Manuscriptoria, tj. těm, kteří skutečně dodávají nebo vbrzku budou dodávat data do Manuscriptoria. Vyžaduje se tudíž autentizace. V uživatelském pohledu je MCan jedním z modulů celkového systému Manuscriptoria, který dovoluje nahrát distribuované komplexní digitální dokumenty a posléze je přesunout do vlastního Manuscriptoria, tzn. zařadit je do jeho stále nabídky. K tomu se ovšem vyžaduje, aby byly zkontrolovány ještě někým dalším, než je jejich tvůrce, kdo je jejich redaktorem, který s konečnou platností schválí jejich zařazení. To je vcelku schůdná cesta, jak rozšiřovat obsah Manuscriptoria na dálku, výhradně operacemi v elektronickém prostředí bez nutnosti fyzického kontaktu. To je další paradigmatická změna, kterou se práce s písemným a dokumentárním dědictvím v síťovém digitálním informačním a komunikačním prostředí liší od práce v prostředí tradičním. Můžeme v tom vidět virtualizaci dokumentu, která odpovídá možnostem MToolu vytvářet nejenom digitální repliky fyzických dokumentů, ale opravdové virtuální dokumenty, jež jako jednotky nemají korelát ve fyzickém prostředí. Je sice zřejmé, že virtuálnost je vlastní informaci vůbec bez ohledu na prostředí, v tomto případě však přestává být pouhou možností a stává se tendencí. Teprve poznáme, jaké to bude mít další následky.

Pátým výstupem je EGE (*ENRICH Garage Engine* – <http://www.dbase.cz:8090/ege/>), což je nástroj pro typizované konverze metadat. Není určen koncovým uživatelům, nýbrž zejména kooperujícím obsahovým partnerům, tj. poskytovatelům dat, kteří se chystají přistoupit k plné spolupráci, tzn. k vlastnímu dodávání dat. EGE dokáže pracovat se standardy TEI, MASTER nebo EAD (*Encoded Archival Description* – <http://www.loc.gov/ead/>), což jsou nejčastější standardy používané při elektronicko-digitálním zpracování písemného a dokumentárního dědictví. Je však třeba poznamenat, že konverzi metadat ze zmíněných standardů do TEI P5 ENRICH Specification nelze považovat za definitivní, protože variabilita, kterou jednotlivé standardy a formáty umožňují, je příliš velká, než aby ji bylo možno spolehlivě zvládnout jedinou predefinovanou konverzí, která nebere ohled na konkrétní použití příslušného standardu a formátu příslušným partnerem. Cílem je spíše testovat konzistenci meta/dat, která vyvářejí jednotliví partneři a která mají být agregována v Manuscriptoriu za účelem zlepšit jejich mapování a dosáhnout tak konverze s co nejmenší ztrátou informace s plným vědomím toho, že bezetrátová konverze je kromě informačně nejhudších dat prakticky nemyslitelná. Kromě tohoto zjevného praktického významu má EGE i méně zjevný smysl konceptuální. Teoreticky totiž připouští reálnou existenci heterogenity dat, snaží se k ní postavit praktickým způsobem a neusiluje o prosazení homogenity dat, jak to zpravidla požaduje tradiční knihovnictví a jak to je obvyklé pro většinu rozšířených a *via facti* závazných tradičních knihovních informačních systémů. Tím, že data jednotlivých partnerů zpřístupňovaná v Manuscriptoriu jsou daty sekundárními, se tudíž nevykonává nepřiměřený tlak na pořizování primárních dat, jež jednotliví partneři pořizují, která jsou v jejich vlastnictví a která jsou zpravidla u nich také uložena, takže partnerské spolupráci se neklade do cesty překážka nadměrných finančních nákladů a nepřiměřené spotřeby času a práce, je tedy pro všechny zúčastněné přijatelnější a snazší. A jelikož na teoretické a metodologické rovině má heterogenní povahu hromadný jev (na rozdíl od jevu jednotlivého), zapadá povolená heterogenita dat dobře mezi ostatní paradigmatické změny, o kterých už byla řeč, a usnadňuje tedy kooperaci v širokém měřítku.

Šestým výstupem je Gaiji Bank (<http://beta.manuscriptorium.com/index.php?q=no-de/3/>), což je databáze nestandardních znaků, s jejímž využitím lze tyto znaky používat v rámci TEI P5 ENRICH Specification. Důležité je, že Gaiji Bank je extenzibilní, v nynější chvíli totiž zahrnuje jen ty nestandardní znaky latinského písma, které jsou používány v nordických jazycích. Nestandardní znaky latinského písma používané v jiných jazycích, jakož i nestandardní znaky v dalších písmech, zejména cyrilském, dosud zařazeny nejsou, protože v rámci trvání projektu ENRICH se nedostávalo ani času, ani pracov-

ních kapacit na provedení všeho, co k tomu bylo zapotřebí. Národní knihovna České republiky předpokládá, že v souvislosti s jejími dalšími aktivitami na poli elektronicko-digitálního zpracování písemného a dokumentárního dědictví bude možno pokračovat v dalším rozšiřování Gaiji Bank, jakkoli se to bude dít především ve spolupráci s partnery Manuscriptoria jak domácími, tak zahraničními. Úvahy na téma nestandardních znaků a budování jejich databáze vedou ovšem dále při konceptualizaci práce s písemným a dokumentárním dědictvím za pomoci informačních a komunikačních technologií. Ukazuje se totiž, že je nutno znovu promýšlet smyslnost i použitelnost transliterace a transkripce. Pokud tradiční důvod pro transliteraci, totiž věrná reprodukce původního zápisu v digitálním prostředí do značné míry padá, poněvadž je k dispozici kopie či replika originálu, v níž je možno – třebaže to ještě není úplně běžné – i navigovat, otázka transkripce se začíná ukazovat v poněkud jiném světle. Pro jazyky, které směřují k fonetickému zápisu, jako je třeba právě čeština, je existence nestandardních znaků výzvou, protože vymizelo povědomí o jejich zvláštní výslovnosti, byla-li jaká, resp. povědomí o zvláštním grafickém úzu, lze li jej definovat. A pro jazyky užívající spíše grafiky řekněme typu etymologického, jsou nestandardní znaky právě takovou výzvou, jenže v opačném pořadí. Po digitální historii, digitální paleografii, digitální kodikologii apod. tak na pole studia písemného a dokumentárního dědictví vstupuje i digitální filologie a komutační lingvistika.

A konečně sedmým výstupem je Systran Multilingual web service with ENRICH profile (<http://enrich.systran.fr/enrich/login.jsp/>), což je první krok na pole multilingvalismu. O jeho úspěšnosti nebo neúspěšnosti nelze v tuto chvíli hovořit, protože spíše než o vyvinutí použitelného a účinného nástroje tu šlo spíše o potřebu alespoň elementárně ohledat celou rozsáhlou problémovou oblast. Systran Multilingual web service with ENRICH profile je v zásadě automatickým překladačem aplikovatelným především na deskriptivní metadata, i když také na plnotextová data. Je zřejmé, že použití na automatický překlad plných textů – třebaže se ukázal jako spolehlivější, a tedy lepší než překlad popisných metadat – nemá přirozené užití, poněvadž typický koncový uživatel Manuscriptoria je mocen jazyků, jimiž jsou psány texty, kterými se zabývá. Totéž samozřejmě platí i o popisných metadatech. Nicméně je třeba si uvědomit, že popisná metadata vystupují ve funkci katalogových záznamů zařazených v databázi evidenčních záznamů, v nichž se vyhledává. Automatický překlad popisných metadat v tom případě nemusí sloužit koncovému uživateli přímo, neboť si takové služby nežádá, nýbrž nepřímo, totiž ve své transformaci do multilingválního vyhledávání. Než se však taková transformace provede, je nutné, aby automatický překlad byl spolehlivý. Od toho je však automatický překlad v nynější chvíli ještě dost daleko: problémem je jak rozdílná slovní zásoba různých historických vývojových stupňů libovolného jazyka, tak rozdílné fonetické a morfologické vlastnosti, tak také rozdílná grafika. Jinými slovy řečeno to znamená, že situace je nesrovnatelně složitější než při multilingválním vyhledávání v současných jazycích, jak je používají vyhledavače typu Google apod. Multilingvální výstup projektu ENRICH svou rudimentárností tedy ukazuje, že ke skutečnému multilingválnímu vyhledávání v písemném a dokumentárním dědictví povede ještě dlouhá cesta.

Rekapitulovat význam projektu ENRICH pro Národní knihovnu České republiky vůbec a pro Manuscriptorium zvláště lze nakonec jen stručně. Pokud jde o obsah, partnerská síť Manuscriptoria se výrazně rozšířila a dostala se na transnacionální evropskou úroveň, přičemž některými kontakty, ať už reálnými (např. Kazachstán) nebo potenciálními (např. Korea), aspiruje na to, aby pronikla až na úroveň globální. Každopádně je nyní Manuscriptorium největší digitální rukopisnou knihovnou na světě a jedním z nejvýznamnějších zdrojů zpřístupňujících písemné a dokumentární dědictví. Pokud jde o stránku technologickou a zejména metodologickou, stalo se Manuscriptorium průkopníkem paradigmatických změn jak při zpracování a zpřístupňování písemného a dokumentárního dědictví, tak v rámci humanitních disciplín všeobecně, zejména však, pokud jde o digitální paleografii, digitální kodikologii, digitální historii, digitální filologii a v neposlední míře i komutační lingvistiku. Koordinování a úspěšné dokončení projektu ENRICH bylo pro Národní knihovnu České republiky i pro Manuscriptorium významným úspěchem, na který by chtěla navázat dalšími aktivitami.