

# ENRICH – PROJEKT EVROPSKÉ INTEGRACE RUKOPISŮ

Zdeněk Uhlíř, Národní knihovna ČR Praha

V loňském roce jsem na patnáctém ročníku konference Knihovny současnosti podal informaci o Manuscriptoriu a jeho snaze být evropskou digitální knihovnou rukopisů.<sup>1</sup> V témže roce jsem na šestnáctém ročníku konference Problematika historických a vzácných knižních fondů Čech, Moravy a Slezska referoval o systematickém zařazování plných textů do Manuscriptoria.<sup>2</sup> A konečně jsem na přelomu jara a léta tohoto roku pro litevský odborný knihovnický časopis Knygotyra připravil stručný výklad o tzv. české digitalizační škole a jejím postupu při zpracování rukopisů dalších historických dokumentů.<sup>3</sup> Na letošní konferenci Knihovny současnosti mohou být tedy poměrně stručný, poněvadž všechny podstatné věci – některé dokonce vícekrát – již zazněly. Ostatně můj příspěvek je tentokrát zcela konkrétní, zaměřuje se na evropský projekt, který Národní knihovna České republiky, resp. tým Manuscriptoria koordinuje.

Národní knihovna České republiky získala v rámci programu eContent-Plus evropský grantový projekt ENRICH (*European Networking Resources and Information Concerning Cultural Heritage*), jehož cílem je vybudovat evropskou digitální knihovnu rukopisů.<sup>4</sup> Je to projekt poměrně rozsáhlý, účastní se ho celkem osmnáct řádných partnerů (*full partners*) z deseti evropských zemí (Česko: Národní knihovna České republiky<sup>5</sup> – AiP Beroun,

---

<sup>1</sup> Viz Uhlíř, Zdeněk: Manuscriptorium na cestě k evropské digitální knihovně. In: Knihovny současnosti 2007. Brno: SDRUK, 2007, s. 136-144. ISBN 978-80-86249-44-5 Prezentace dostupná na URL: <http://www.sdruk.cz/sec/2007/sbornik/0-08.pdf> (8.8.2008).

<sup>2</sup> Viz Uhlíř, Zdeněk: Manuscriptorium: evropská digitální knihovna rukopisů s plnými texty. In: Problematika historických a vzácných knižních fondů Čech, Moravy Slezska 2007. Olomouc: VKOL-SDRUK, 2008, s. 103-108. ISBN 978-80-7053-276-8, 978-80-86249-47-6.

<sup>3</sup> Viz Uhlíř, Zdeněk: Digitization is not only Making Images. In: Knygotyra (v tisku). ISSN 0204.2061. Srv. <http://www.leidykla.eu/mokslo-darbai/knygotyra/> (8.8.2008).

<sup>4</sup> Viz ENRICH: European Networking Resources and Information concerning Cultural Heritage. Dostupné z URL <http://enrich.manuscriptorium.com/> (8.8.2008).

<sup>5</sup> Viz <http://nkp.cz/> (8.8.2008).

s.r.o.<sup>6</sup> – Crocc Czech, a.s., Praha;<sup>7</sup> Velká Británie: Oxford University Computing Services;<sup>8</sup> Francie: SYSTRAN, S.A., Paris;<sup>9</sup> Španělsko: Biblioteca nacional de España, Madrid;<sup>10</sup> Itálie: Biblioteca nazionale centrale di Firenze<sup>11</sup> – Centro per la comunicazione e l'integrazione dei media, Firenze;<sup>12</sup> Rakousko: Diözesanarchiv St. Pölten;<sup>13</sup> Německo: Die Historisch Kulturwissenschaftliche Informationsverarbeitung, Universität zu Köln;<sup>14</sup> Maďarsko: Budapesti Műszaki és Gazdaságtudományi Egyetem;<sup>15</sup> Polsko: Biblioteka Uniwersytecka we Wrocławiu<sup>16</sup> – Poznańskie Centrum Superkomputerowo Sieciowe;<sup>17</sup> Litva: Vilniaus universiteto biblioteka<sup>18</sup> – Matematikos ir informatikos institutas, Vilnius;<sup>19</sup> Dánsko: Nordisk Forskningsinstitut, Københavns Universitet;<sup>20</sup> Island: Stornuj Árna Magnússonar, Reykjavík<sup>21</sup> – Landsbókasafn Íslands-Háskólabókasafn, Reykjavík<sup>22</sup>). Vedle toho se k projektu řadí i přidružení partneri (*associated partners*), kteří oficiálně vyjádřili zájem podílet se na uskutečňování cílů projektu na vlastní náklady; značná část přidružených partnerů chápe svou účast spíše platonicky a zatím nevyvíjí žádnou činnost, jsou však i takoví, kteří by mohli být příkladem pro partnery řádné, zejména Universitätsbibliothek Heidelberg<sup>23</sup> a Biblioteca Nationala României, București.<sup>24</sup>

Cílem projektu ENRICH je integrovat do digitální knihovny Národní knihovny České republiky Manuscriptorium co nejvíce co nejbohatších dalších zdrojů týkajících se evropského písemného a zejména rukopisného kulturního dědictví. Nyní digitální knihovna Manuscriptorium obsahuje více než 3 000 plně digitalizovaných dokumentů, tj. asi 1 100 000 digitálních ob-

---

<sup>6</sup> Viz <http://www.aipberoun.cz/> (8.8.2008).

<sup>7</sup> Viz <http://crossczech.cz/> (8.8.2008).

<sup>8</sup> Viz <http://www.oucs.ox.ac.uk/> (8.8.2008).

<sup>9</sup> Viz <http://www.systran.fr/> (8.8.2008).

<sup>10</sup> Viz <http://www.bne.es/> (8.8.2008).

<sup>11</sup> Viz <http://www.bncf.firenze.sbn.it/> (8.8.2008).

<sup>12</sup> Viz <http://lci.det.unifi.it/> (8.8.2008).

<sup>13</sup> Viz <http://www.dsp.at/dasp/> (8.8.2008).

<sup>14</sup> Viz <http://www.hki.uni-koeln.de/> (8.8.2008).

<sup>15</sup> Viz <http://portal.bme.hu/default.aspx> (8.8.2008).

<sup>16</sup> Viz <http://www.bu.uni.wroc.pl/> (8.8.2008).

<sup>17</sup> Viz <http://www.man.poznan.pl/pcss/public/main/index.html> (8.8.2008).

<sup>18</sup> Viz <http://www.mb.vu.lt/> (8.8.2008).

<sup>19</sup> Viz <http://www.mii.lt/> (8.8.2008).

<sup>20</sup> Viz <http://nfi.ku.dk/> (8.8.2008).

<sup>21</sup> Viz <http://www.arnastofnun.is/> (8.8.2008).

<sup>22</sup> Viz <http://www.bok.hi.is/> (8.8.2008).

<sup>23</sup> Viz <http://www.ub.uni-heidelberg.de/> (8.8.2008).

<sup>24</sup> Viz <http://www.bibnat.ro/> (8.8.2008).

razů, každý v pěti různých kvalitách (galerie, náhled, nižší, vyšší, černo-bílá optimalizace). V době ukončení projektu ENRICH předpokládáme, že tento počet se zhruba zpětinasobí, tzn. že Manuscriptorium jakožto zdroj bude obsahovat asi 5 000 000 digitálních obrazů. Tyto obrazy ovšem nebudou centralizovány v databázi Manuscriptoria. Centralizována budou jenom metadata (zejména popisná a strukturální, avšak také technická) a obrazy budou distribuovány v síťovém prostoru (*cyberspace*), řečeno jinak zůstanou u svých tvůrců, resp. vlastníků. Jak už jsem se zmínil, nespokojujeme se s konsorciem řádných partnerů, ale intenzivně se snažíme vyhledávat i partnery přidružené a s tím počítáme po skončení projektu ENRICH. K tomu je určena osmá pracovní oblast (*workpackage*), kterou je propagace a získávání dalších činných partnerů mimo základní okruh partnerů řádných (*dissemination*). Propagace a rozšiřování je ovšem typickou, ba povinnou součástí všech evropských projektů, právě tak jako oblast organizace a managementu (v případě projektu ENRICH *workpackage number 1*), bez které se přirozeně neobejde žádná týmová a v jistém smyslu ostatně ani individuální práce.

Specifické pro projekt ENRICH jsou pracovní oblasti mezi organizací a managementem na jedné a propagací a získáváním partnerů na druhé straně. Takže je tu pracovní oblast druhá (*workpackage number 2*), tj. příprava pro implementaci systému a obohacení obsahu (*preparation for system implementation and content enhancement*), třetí (*workpackage number 3*), standardizace shromážděných metadat tj. (*standardization of shared metadata*), čtvrtá (*workpackage number 4*), tj. personalizace pro koncové uživatele (*user personalization*), pátá (*workpackage number 5*), tj. personalizace pro přispěvatele (*personalization for contributors*), šestá (*workpackage number 6*), tj. mnohojazyčné a výkonné uživatelsky vřidné zpřístupnění (*multilingual an user friendly sophisticated access*) a sedmá (*workpackage number 7*), tj. hodnocení, testování a ověřování (*evaluation, testing and validation*).

Oblast přípravy pro implementaci systému a obohacení obsahu spočívá zejména za prvé v průzkumu kvantity a kvality dat poskytovaných jednotlivými partnery, jehož výstupem je zpráva podávající přehled výsledků a jejich interpretaci (*Survey results and their interpretation*), a za druhé v konkrétním popisu materiálu poskytovaného partnery a jeho výstupem je zpráva spočívající v popisu standardů užívaných jednotlivými partnery, určení principů spolupráce, jakož i datových a metadatových standardů (*Description of the standards used by the partners, definition of collaboration principles, data and metadata standards*). Cílem této skupiny činností je připravit všechny potřebné podklady pro reálnou integraci dat a pro jejich efektivní zpřístupnění v jediném uživatelsky vřidném rozhraní.

Oblast standardizace shromážděných metadat spočívá zejména v koordinaci se standardem TEI,<sup>25</sup> přičemž výstupem je zpráva podávající specifikaci standardu TEI P5 (*Revised TEI conformant specification*), v přípravě prakticky použitelných materiálů pro přispěvatele, kde je výstupem zpráva v podobě dokumentace a cvičných materiálů (*Documentation and training materials for use with specification*), v podpoře konverzí dat, resp. metadat jednotlivých partnerů, kde je výstupem zpráva o konvertorech a podobných nástrojích (*Report on development and validation of migration tools*) a konečně v důrazu na interoperabilitu, kde je výstupem zpráva o souvztažnosti standardů METS<sup>26</sup> na jedné a TEI na druhé straně, jakož i o využívání kódování v UNICODE i jiného než UNICODE (*Report on METS/TEI interoperability, best practice with respect to handling of Unicode and non-Unicode data in Manuscriptorium and P5 conversion techniques*). Cílem této skupiny činností je podpora hladkých konverzí a transformací dat, resp. metadat a bezproblémové systémové interoperability.

Oblast personalizace pro koncové uživatele spočívá zejména v přípravě podkladů pro vytváření virtuálních kolekcí a virtuálních dokumentů, kde je výstupem zpráva definující požadavky na vytváření personálních digitálních knihoven (*Definition of requirements for the creation of personalised virtual digital libraries*), v průzkumu chování uživatelů Manuscriptoria při vyhledávání, kde je výstupem shrnující zpráva o uživatelských zvyklostech (*Report on search behaviour of Manuscriptorium users*), v hodnocení vytvořených nástrojů pro koncové uživatele, kde je výstupem zpráva o jejich pilotní implementaci (*Report on pilot implementation of tools for creation of virtual documents by researchers*) a konečně ve zkoumání typických modelových způsobů vyhledávání, kde je výstupem zpráva definující tyto modely (*Report on definition of typical model searches and their implementation*). Cílem této skupiny činností je podpora věnovaná vytvoření vlivného a zároveň účinného prostředí pro koncové uživatele.

Oblast personalizace pro přispěvatele spočívá zejména v praktickém určení postupu při integraci dat, resp. metadat, kde výstupem je zpráva definující základní podmínky pro integraci hromadných dat v Manuscriptoriu (*Definition of basic conditions for sharing of large data sets in the frame of Manuscriptorium*), v pilotním zařazení prvních partnerských zdrojů, kde výstupem je zpráva o plné integraci a konečném zpřístupnění dat vybraných partnerů v Manuscriptoriu (*Report on pilot full integration and publication of selected partner's metadata in Manuscriptorium*) a v zařazení dokumentů těch partnerů, kteří poskytují data externě, tzn. ze vzdáleného serveru, kde

---

<sup>25</sup> Viz <http://www.tei-c.org/> (8.8.2008).

<sup>26</sup> Viz <http://www.loc.gov/standards/mets/> (8.8.2008).

výstupem je zpráva o plné integraci a konečném zpřístupnění v Manuscriptoriu metadat vybraných partnerů a externě uložených dat (*Report on pilot full integration and publication of selected partner's metadata and externally stored data in Manuscriptorium*). Cíl této skupiny činností je pro projekt ENRICH a pro Manuscriptorium jakožto evropskou digitální knihovnu rukopisů zcela zásadní, protože právě zde jsou jak výsledek veškerých činností souvisejících s integrací, tak obsah sám, tj. data a metadata všech řádných i přidružených partnerů, zpřístupněny koncovému uživateli.

Oblast mnohojazyčného a výkonného uživatelsky vlídného zpřístupnění spočívá v průzkumu metadat poskytovaných jednotlivými řádnými partnery projektu ENRICH z hlediska multilingualismu, kde výstupem je analýza korpusu (*ENRICH Corpus Analysis report*), v implementaci multilingválního překladového stroje, kde výstupem je zpráva o personalizovaném překladovém rozhraní (*Personalised Translation Interface delivery report*), v určení elementů ve struktuře metadat, které mají být automaticky zařazeny k překladu, kde výstupem je zpráva popisující schematicky tyto elementy (*ENRICH Translation Stylesheet delivery report*), a konečně v pilotní implementaci a testování vyhledávání na základě ontologií, kde výstupem je zpráva o implementaci systému ontologií vycházejících z dřívějšího evropského projektu VICODI<sup>27</sup> (*Vicodi Ontologies implementation report*). Cílem této skupiny činností je využití multilingualismu nejenom v běžném významu různojazyčných lokalizací uživatelského rozhraní, ale také v hlubším významu rozšíření vyhledávacích funkcí informačního systému.

Oblast hodnocení testování a ověřování spočívá zvláště v přehledu dostupných metodologií pro testování a ověřování nástrojů, kde výstupem je zpráva definující principy hodnocení technických aplikací vzniklých v rámci projektu (*Methodology for Testing and Validation of e-Applications*), a v praktickém hodnocení výsledků a výstupů projektu, kde výstupem je zpráva hodnotící základní technické nástroje, které jsou výstupem projektu (*The Evaluation Report*). Cílem této skupiny činností je důsledné hodnocení výsledků a výstupů projektu ENRICH jak v ohledu obsahu, tj. kvality dat a metadat, která podléhají integraci, tak v ohledu informačního systému jako takového, resp. jeho jednotlivých nástrojů vzniklých v průběhu projektů, aby pro Manuscriptorium jakožto zdroj byla zajištěna nejenom velká kvantita integrovaných dat a metadat, ale také jejich vysoká kvalita.

Vyvoření základní infrastruktury pro zpřístupnění evropského rukopisného kulturního dědictví je jistě velký úkol. A jestliže v něm konsorcium projektu ENRICH uspěje, bude to jistě významný příspěvek k budování di-

---

<sup>27</sup> Viz <http://www.vicodi.org/> (8.8.2008).

gitální Evropy všeobecně. Nicméně vytvoření této základní infrastruktury je jen jedna stránka věci. Druhou stránkou je vytvoření opravdového uživatelského prostředí, virtuálního badatelského prostředí pro práci s historickými fondy, které by sdružilo jak svět paměťových institucí (knihoven, archivů, muzeí apod.), tak svět univerzitní a akademických institutů, tak také svět všech více či méně kultivovaných individuálních uživatelů a všechny tyto dílčí světy společně vybavilo co nejrozmanitějšími nástroji pro využívání historických dokumentů, jejich kolekcí nebo naopak částí v tom či onom směru. To je konečným cílem Národní knihovny České republiky a týmu Manuscriptoria. Proto se také Národní knihovna České republiky zapojila do konsorcia připravujícího evropský grantový projekt ANCHISAE.<sup>28</sup> Ale to je zatím jen budoucnost.

---

<sup>28</sup> Viz Edmond, Jennifer: The Anchisae Project. Dostupné z URL: [ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/20080418-infoday-edmond\\_en.pdf](ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/20080418-infoday-edmond_en.pdf) (8.8.2008).