

Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	Mgr. Jindřich Marek, Ph.D.
Pracoviště – dle organizační struktury	1.5.1 ORST
Pracoviště – zařazení	Oddělení rukopisů a starých tisků – odborný pracovník
Důvod cesty	účast na výročním zasedání konsorcia TEI
Místo – město	Evanston, IL
Místo – země	Spojené státy americké
Datum (od-do)	21.–27. 10. 2014
Podrobný časový harmonogram	<p>21. 10. 8.50 odlet z Letiště Václava Havla Praha 14.10 přílet na Chicago O'Hare International Airport</p> <p>22.-25. 10. účast na výročním zasedání konsorcia TEI, program: http://tei.northwestern.edu/program/, srov. též https://twitter.com/hashtag/teiconf2014?f=realtime</p> <p>v tom:</p> <p>23. 10. účast na setkání zájmové skupiny k rukopisům; stránky pracovní skupiny: http://wiki.tei-c.org/index.php/SIG:MSS účast na setkání zájmové skupiny ke knihovnám; zápis: http://wiki.tei-c.org/index.php/Minutes_from_October_23,_2014</p> <p>24. 10. účast na setkání členů konsorcia</p> <p>25. 10. účast na celodenním workshopu věnovaném odvozování schémat TEI; program: http://tei.it.ox.ac.uk/Talks/2014-10-odds/programme.xml</p> <p>26. 10. 17.25 odlet z Chicago O'Hare International Airport</p> <p>27. 10. 13.30 přílet na Letiště Václava Havla Praha</p>
Spolucestující z NK	–
Finanční zajištění	VaV (0132)
Cíle cesty	<ul style="list-style-type: none"> • Účast na výročním zasedání konsorcia TEI. • Sledování nových trendů v TEI a XML obecně. • Zapojení do diskuse na setkání zájmových skupin k rukopisům a ke knihovnám. • Účast na workshopu věnovanému schématu TEI ODD odvozování schémat.
Plnění cílů cesty (konkrétně)	<p>Cíle cesty byly splněny.</p> <p>Výroční konference představila nové trendy v TEI. Mezi ně patří na jednom z předních míst snaha o sjednocení kódování v jednotlivých textových korpusech kvůli udržitelnosti projektů. O to usiluje</p>

Zpráva je pracovníkem do mezinárodního oddělení předložena nejpozději při vyúčtování cesty do 2 týdnů po jejím ukončení. Bez cestovní zprávy nebude provedeno vyúčtování. Při výjezdu více pracovníků na tutéž služební cestu s týmž programem lze odevzdat společnou cestovní zprávu.

schéma TEI Simple (o něm níže) a projekty TAPAS a EEBO-TCP (v prvním případě se jedná o úložiště pro texty podle schématu TEI s integrovanými nástroji, ve druhém případě o korpus textů, na jehož udržování se podílí zejména univerzita v Oxfordu). Základem je nespolehat se na grafickou podobu výstupu, protože jinak dochází zpravidla ke zneužívání tagů a kódování je příliš zaměřené na výstup. Protože je obtížné validovat přizpůsobená TEI ODD schémata, zejména pokud byly přejmenovány elementy, padl návrh na výlučné využívání tříd existujících elementů pro vytváření nových.

Schéma TEI Simple, jež představil James Cummings, je zaměřeno především na samotný text, který je dokumentem zprostředkovan. Cíleno je hlavně na moderní tištěné knihy. Nejedná se o snahu markup maximálně zjednodušit, ale struktura dokumentu má být při relativní jednoduchosti co nejvíce efektivní. Vytvoření schématu TEI Simple je založeno na analýze užití markupu ve velkých textových korpusech, jež ukázala, že malé množství elementů pokryje naprostou většinu užití. Se schématem jsou spojena i určitá doporučení týkající se užívání atributů a jejich hodnot. Seznam elementů a atributů dosud není uzavřen. Cílem je standardizace, normalizace a zajištění kompatibility. Do blízké budoucnosti je plánováno namapování schématu do RDF.

Se schématem TEI Simple souvisí také připravovaný vývoj nové verze standardu označené P6, protože vychází ze stejných trendů a potřeb (srov. <http://wiki.tei-c.org/index.php/P6-dev>).

Ke kořenům TEI se obrátilo zamyšlení nad charakteristikou tohoto heterogenního formátu a technologických principů, na nichž je založen. Sebastian Rahtz představil atraktivní vizualizaci schématu a připomněl, že definice formátu sestávají z 39 TEI XML souborů, z toho 25 jsou podklady pro jednotlivé kapitoly *TEI Guidelines*.

Srov. <http://tei.oucs.ox.ac.uk/Talks/2014-10-teimm-visualization/circle-tree.html>.

Srov. též <http://tei.oucs.ox.ac.uk/Talks/2014-10-teimm-visualization/teisimple-force2.html>.

Stejný mluvčí se zamyslel nad konverzí ze starších formátů užívaných pro EEBO-TCP do formátů TEI P5 a epub (dostupné v Apple Store). Tyto úvahy jsou důležité i kvůli tomu, že v TEI XML není jasně definovaná hierarchie elementů a není možné reprezentovat ontologii. Díky referenci k jiným formátům je možné zjednodušení některých struktur (použití výstižnějšího elementu v P5) a v některých případech byly dokonce odhaleny nedostatky v návrhu schématu TEI.

Trendem je také stále rozšířenější užívání TEI i ve státní správě (např. US Department of State, též na utajované materiály), mj. pro vydávání historických dokumentů (v uvedeném příkladu dokumentů

k zahraniční politice USA) v elektronické i tištěné podobě.

Ze zajímavostí důležitých pro zpracování historických fondů byl představen program na rozpoznávání (OCR) pro raněnovověké texty: <http://emop.tamu.edu/>. Tyto technologie ve velmi blízké budoucnosti zajistí masivní zpřístupnění textů ze starých tisků a přitom zprostředkují také úplné informace z titulních listů, jež dnes katalogizátoři zpravidla v úplnosti přepisují, aniž by se při této časově náročné a vlastně zbytečné práci pokusili o zkrácení a rektifikaci titulu díla.

Setkání členů konsorcia přineslo kromě výsledků voleb do statutárních orgánů TEI také informace o struktuře členství v konsorciu TEI a o časopisu TEI Journal (<http://journal.tei-c.org/journal/index>).

Zájmová skupina k rukopisům, jež se sešla za předsednictví Dot Porter, se věnovala diskusi o zápisu a grafické reprezentaci kolace složek v rukopisech a o popisu knižního bloku zejména v kontextu kompozitních středověkých rukopisů. Na přetřes přišla i knižní vazba.

Dalším tématem, odvozeným spíše od ediční techniky, byla transkripce textů (srov. návrh <http://bit.ly/1DEbmN6>). Elena Pierazzo zpochybnila XML jako jazyk vhodný pro vyjádření paralelní segmentace textů. Byla řešena také otázka kolace textů v rukopisech, způsob jejího provádění a reprezentace v XML. Torsten Schassan se přimlouval za transkripce jednotlivých rukopisů a následnou automatickou kolaci pomocí nástrojů. Elena Pierazzo zdůraznila, že neexistují nástroje, které by byly schopné reprezentovat v TEI výsledky automatické kolace (elementy pro kritický aparát nejsou pro tento účel vhodné). Dalším limitujícím faktorem je, že při srovnávání textů se srovnávají vždy vzájemně pouze dva rukopisy; srovnat více rukopisů najednou není možné. Diskuse se dále vyvinula k otázce, co je to edice a jaký pro ni vytvořit datový model.

Závěrem bylo, že pojetí textu v papyrologii a medievistice (v těchto oborech pracují všichni přítomní) se různí a že bude třeba zformovat pracovní skupinu pro kritický aparát, jež bude důsledně rozlišovat mezi kolací-komparací textů a reprezentací kritického aparátu.

Představil jsem záměr Národní knihovny zpracovat v následujícím plánovacím období v rámci základního výzkumu provenience středověkých knih, k němuž už bylo nakročeno první fází zpracování knižních vazeb. Tento záměr byl hodnocen velmi pozitivně, také s očekáváním případných budoucích úprav schématu TEI pro popis rukopisů, aby do něho bylo možné vkládat kvalifikované informace (ve smyslu strukturační dat) o exemplářových znacích středověkých knih, tedy proveniencích a knižních vazbách.



	<p>Na setkání zájmové skupiny ke knihovnám se mnoho hovořilo o konverzi mezi různými metadatovými formáty, zejména z TEI do běžných knihovnických formátů poskytovaných v knihovnických systémech navenek protokolem OAI-PMH. Ačkoliv diskuse byla zaměřena především na knihovny v USA, poreferoval jsem o výsledcích dosažených na tomto poli v oblasti historických fondů Národní knihovnou. Byla diskutována také digitalizace v rámci projektu Google Books, jež je v USA i u nás aktuální. Přitom se ukázalo, že podmínky ve smlouvách se společností Google jsou i pro velké a významné severoamerické knihovny výrazně tvrdší, než jaké se podařilo dojednat Národní knihovně.</p> <p>Workshop zaměřený na odvozování schémat pomocí TEI ODD, jež je důležité také kvůli sdílení schématu a dokumentace, poskytl praktický návod na odvozování přizpůsobených schémat, který bude možno v budoucnosti využít například při vytváření schématu zohledňujícího zápis exemplářových znaků středověkých knih pro potřeby základního výzkumu v Národní knihovně.</p>	
Program a další podrobnější informace	–	
Přivezené materiály	–	
Datum předložení zprávy	10. 11. 2014	
Podpis předkladatele zprávy		
Podpis nadřízeného	Datum:	Podpis:
Vloženo na Intranet	Datum:	Podpis:
Přijato v mezinárodním oddělení	Datum:	Podpis:

Zpráva je pracovníkem do mezinárodního oddělení předložena nejpozději při vyúčtování cesty do 2 týdnů po jejím ukončení. Bez cestovní zprávy nebude provedeno vyúčtování. Při výjezdu více pracovníků na tutéž služební cestu s týmž programem lze odevzdat společnou cestovní zprávu.