

Národní kulturní dědictví v digitálním světě

Ing. Luděk TICHÝ

Národní knihovna České republiky

ludek.tichy@nkp.cz

INFORUM 2014: 20. ročník konference o profesionálních informačních zdrojích

Praha, 27. -28. 5. 2014

Anotace : Hlavní faktory problematiky zpřístupnění digitálního obrazového obsahu a jejich dopad. Velké objemy dat, dostupnost, finanční náročnost ve vazbě na kapacity ukládání a infrastruktury. Vazba na téma BigData, zabezpečení a autorský zákon, podpora vzdělanosti veřejnosti a systémová podpora vzdělávání.

Na základě zkušeností z realizace projektu NDK a následně masové resp. velkoobjemové digitalizace monografií a periodik, usilujeme o spolupráci s dalšími paměťovými a uměleckými institucemi v rámci resortu kultury na téma digitalizace národního kulturního dědictví. S pohledu historie se kulturním dědictvím stává každý předmět digitalizace, který nějakým způsobem vypovídá o historii či o charakteru a průběhu doby. Z tohoto hlediska vše počínaje elektronickou poštou, texty, fotografie, multimédia, hry atd. lze považovat za předmět kulturního dědictví. Toto tvrzení napovídá, že se jedná o velké objemy dat a není v silách současných institucí tato data uchovávat (z technických, personálních, finančních a jiných důvodů). Pokud ovšem vezmeme v úvahu reálné možnosti současné společnosti, je určitě možno stanovit dosažitelné cíle a proveditelný postup.

Jako ve většině případů i zde je nutno vytvořit například „národní strategii digitalizace kulturního dědictví“. Nemusí nutně jít o technicky či odborně vyspělý materiál, naopak musí vzniknout laickou veřejností čitelný a pochopitelný dokument, který definuje především předmět digitalizace s ohledem na konkrétní obor a k němu stanovit definici způsobu zachycení objektu a formát uložení do LTP (long term preservation). Například víme, že je nutno digitálně zpracovat a uložit objekty či exponáty muzeí, ale již není jasné, jestli se tak má stát prostřednictvím fotografie, 3D obrazu nebo videa a o formátu nemluvě. Další problém představuje standard metadat, kde je nutno stanovit metadatovou specifikaci pro resort tak, aby získaná data bylo možno bez problémů ukládat do rezortního úložiště a následně zajistit jejich dostupnost i za řádově desítky let. Strategie musí dále definovat procesy a to nejen jak data pořizovat a ukládat, ale i jak o ně pečovat po neomezeně dlouhou dobu, jak je kontrolovat, jak udržovat konzistenci a čitelnost a v neposlední řadě jak postupovat v případě, že se změní například užitý formát. Velmi důležité je, aby dokument obsahoval ekonomickou rozvahu v čase, jednoduše harmonogram financování, protože v zásadě není nutno digitalizovat všechno a hned ale postupovat tak, aby se prioritně digitalizovalo to co je důležité např. z důvodu ochrany objektu před erozí, zcizením, poškozením atd. Ve své podstatě se jedná o projektovou fiši resp. projektový záměr, tak aby byl jasně definován cíl, předmět, finanční a časový rámec a bylo možno začít myšlenku digitalizace realizovat.

Jelikož v oboru knihovnictví je digitalizace a časově garantované ukládání jako disciplína mnohem dále než v jiných resortních oborech, mohu si dovolit již dnes předpovídat úskalí, na která proces digitalizace kulturního dědictví bude narážet. Pochopitelně jsou to v konečném důsledku ekonomické aspekty, neboť kultura nepatří mezi výrazné „business“ aktivity

(přesněji památky apod.) a je zde nutný kontinuální příděl státních prostředků, kde se zdroj také intenzivně zmenšuje. Toto ovšem neznamená, že je digitalizace v této oblasti nemožná a nerealizovatelná. Následně se pokusím navrhnout postup v oblasti knihovnictví a tento je jako takový aplikovatelný i v ostatních resortních oblastech jako je architektura, výtvarné umění, filmová tvorba, živé umění obecně atd.

Současným asi nejžhavějším tématem a fenoménem ICT je pojem BigData. Jedná se o logický důsledek vzniku kybernetického světa, snahu vytvořit obdobu reálného prostředí a to pro osobní i pracovní život, pro práci, zábavu vzdělávání atd. Tato snaha je charakterizována masivní digitalizací veškerých informací tzn. převodem jejich „analogové“ či materiální podoby do „digitální“ či virtuální podoby. Veškeré komerční, veřejnoprávní a ostatní subjekty se intenzivně zabývají buď digitalizací samotnou, anebo ji plánují. Logicky takto vznikají velké objemy dat, které je potřeba nejen uložit, ale následně i zpracovat. Ukládání dat již delší dobu není problém, co ovšem je problém, je dostupnost. Dostupnost jako taková je do jisté míry řešitelná běžnými nástroji, jako je zvyšování výkonu, kapacity, šířkou pásma apod. Problém ovšem nastal vznikem potřeby velké objemy dat zpracovávat. Zde je rychlost zpracování limitována např. již omezením rychlostí čtení točivého media, jako je HDD, omezením průtoku dat na sběrnici atd. Toto je nějakým způsobem řešitelné, ovšem v kombinaci s problémy na aplikační vrstvě se dostáváme k tématu BigData, které se začíná zásadním způsobem lišit od „klasického“ zpracování dat. Většinou předpokládá práci s daty z různých zdrojů, v různých formátech a v různém rozsahu. Takto získaná data chceme analyzovat a získávat tak další informace, ovšem za použití běžných nástrojů toto není proveditelné v přijatelném čase a proto technologičtí lídři v oblasti ukládání dnes usilovně vyvíjí nástroje k práci s „velkými daty“ a to jak na hw, tak i sw úrovni. Pro oblast kultury je toto téma obzvláště zapeklité, neboť se jedná v zásadě o obrazová či multimediální data a ty jsou velká sama o sobě na úrovni souboru, natož pak v množině souborů. Zatím v oblasti zpracování obrazových či multimediálních dat není k dispozici příliš mnoho ověřených či dostupných řešení nebo platforem. Zatím se rýsuje možnost ve využití open source framework Hadoop, kde dosavadní výsledky z užití ukazují, že se jedná o vhodně nastavenou cestu k zpracování velkých dat.

Dalším velkým tématem je CLOUD. Jednoduše řečeno, virtuální shluk počítačů velké kapacity do jednoho dostupného „mraku výkonu“ a možnost sdílení tohoto výkonu velkým množstvím uživatelů. Tento nový způsob využívání výpočetního výkonu je pro nás obzvláště zajímavý právě z důvodu, že vznikl z čistě ekonomického pohledu na provozování ICT služeb. Provozovatel Cloudu shromažďuje velký výkon spolu s aplikacemi, které pak může nabídnout velkému množství uživatelů tak, aby sdíleli pouze výkon a aplikace, který využijí s odpovídající cenovou relací. Paměťové instituce či obecně příspěvkové organizace jsou velmi často v situaci, kdy šetří na nákupu VT a už vůbec si nemohou dovolit platit kompetentní personál k provozu většího množství specializovaných aplikací a proto by bylo velmi vhodné nabídnout tuto variantu řešení ICT služeb v rámci resortu. Z hlediska resortu je možno realizovat provoz cloudu ekonomicky a efektivně s velkým ekonomickým přínosem pro zřizované organizace. Dalším benefitem je možnost cloud využít i směrem k veřejnosti či ostatním resortům pro sdílení digitalizovaného obsahu.

Pro zmíněnou strategii považujeme uvedená témata za možný rámec pro řešení následné digitalizace. Technologie na pořízení dat je dostupná, řešení ukládání do LTP považujeme v tento okamžik za technologicky vyřešené a je potřeba zvládnout v následujícím čase pouze procesní stránku věci, zůstává nám otázka řešení platformy pro zpřístupnění tedy BigData.

Zpřístupnění samotné lze realizovat jako resortní cloud, což výrazně zvýší dostupnost dat a také pořizování dat z více zdrojů. Tímto je podpořena myšlenka vzniku resortního datového centra, kde lze tyto technologie umístit a provozovat. Posledním zásadním tématem je autorský zákon. Ač to není na první pohled patrné, vznik centralizovaného řešení s tímto tématem úzce souvisí. Pokud je cílem centrální zdroj digitalizovaných dat je třeba zajistit, aby všechny knihovny byly oprávněny pracovat se všemi daty zde uloženými. Toto je možno jen v případě, že autorský zákon bude obsahovat jednoduchá pravidla pro zpřístupnění digitalizovaného či digitálního obsahu ve veřejných knihovnách a za účelem vzdělávání. Nemohu v současnosti předjímat, jak by tato pravidla měla vypadat, nicméně současná legislativa je toto umožňuje za takových podmínek, že z technického hlediska je řešení neefektivní (komplikované a drahé). Právě vznik vhodné legislativy by umožnil opravdu produktivní využití digitalizovaného kulturního dědictví pro účely výuky. Pracujeme s poznáním, že veškerá vzdělanost vychází z literatury resp. z tištěné podoby. Pokud bychom mohli obsah poskytnutý Národní digitální knihovnou obohatit o další kulturní obsah, máme k dispozici unikátní soubor informací pro vzdělávání. Lze je využít pro tvorbu výukových programů, prezentací, vědeckou činnost atd. Opět je zde vysoká přidaná hodnota nejen v podobě centralizovaného obsahu, ale i dále vytváření dalšího obsahu kombinací a zpracováním primárního obsahu.

Závěrem lze uvést zásadní zkušenost s digitalizací. Čím více digitalizujeme, tím více je co digitalizovat!