

# Kvalitní studie k problémům dlouhodobé archivace digitálních dokumentů

Eva Bratková, Ústav informačních studií a knihovnictví FF UK v Praze

E-mail: [Eva.Bratkova@ff.cuni.cz](mailto:Eva.Bratkova@ff.cuni.cz)

## Citace původního článku (recenze):

BRATKOVÁ, Eva. Kvalitní studie k problémům dlouhodobé archivace digitálních dokumentů. *Ikaros* [online]. 2011, roč. 15, č. 4 [cit. 2011-05-31]. Dostupný z WWW: <<http://www.ikaros.cz/node/6754>>. ISSN 1212-5075.

## English title:

First-rate study to issues of digital documents long-term preservation

## Abstract:

The review evaluates the structure, content and formal qualities of the new book titled „Long-term preservation of digital documents“ (published in Czech by the National Library of the Czech Republic, 2010, ISBN 978-80-7050-588-5). The book is based on the text of author's master thesis. The reviewed work aims at the conceptualisation of digital preservation in memory institutions. Its author presents new discipline known as digital preservation. The book focuses on an analysis of all major risks associated with long-term digital heritage preservation. The book also contains an overview of the basic assumptions and objectives of digital preservation. The largest section of the book is devoted to practical outcomes of digital preservation, i.e. preservation actions to ensure digital heritage sustainability. In conclusion the reviewed book presents an analysis of digital preservation development as well as perspectives for the future of digital heritage sustainability.

**Klíčová slova:** elektronické zdroje, digitální objekty, digitální dědictví, digitální archivace, digitální archivy, Otevřený archivní informační systém (OAIS), ISO 14721:2003

**Keywords:** electronic resources, digital objects, digital heritage, digital preservation, digital archive, Open Archival Information System (OAIS), ISO 14721:2003

## Recenzovaná publikace:

**CUBR, Ladislav. *Dlouhodobá ochrana digitálních dokumentů*. 1. vyd. Praha : Národní knihovna České republiky, 2010. 154 s. ISBN 978-80- 7050-588-5 (brož.).**

Nestává se než výjimečně, aby absolvent magisterského studia oboru Informačních studií a knihovnictví převedl v relativně krátké lhůtě po obhajobě svoji diplomovou práci do podoby



knižní publikace. Autor recenzované knihy Ladislav Cubr byl v tomto směru více než úspěšný. Národní knihovna České republiky mu koncem minulého roku vydala knižní publikaci (v brožované vazbě), jejíž základní obsah se zrodil za příznivých okolností. Text byl původně připravován v roce 2009 jako diplomová práce pod názvem „Strategie ochrany digitálních dokumentů“ (tuto skutečnost autor uvádí v předmluvě

publikované práce na s. 8). Autor ale již v té době také působil v Národní knihovně ČR v oddělení, které se zabývá výzkumem dlouhodobé archivace dokumentů a které zajišťuje

budování systému Národní digitální knihovny (NDK). Bylo proto logické a také vítáno, že autor spojil téma diplomové práce s vlastní výzkumnou činností na zmíněném pracovišti. (Jak vyplývá až z informací uvedených na dodatečně vlepeném štítku na rubu titulní strany recenzované publikace, šlo o příspěvek k řešení výzkumného záměru MK00002322102). Diplomová práce byla obhájena jako vynikající a autorovi bylo doporučeno, aby text publikoval a aby ho zároveň rozšířil do podoby rigorózní práce. Obě doporučení autor realizoval. V dubnu 2010 rigorózní práci s pozměněným názvem „Trvalá udržitelnost digitálního dědictví“ taktéž úspěšně obhájil. Text pravděpodobně poté po dalších úpravách připravil do tisku.

Recenzovaná publikace je věnována naléhavé otázce dnešních dnů – složité problematice dlouhodobé ochrany a zpřístupňování digitálních dokumentů. Jak autor uvádí v předmluvě, chtěl, kromě jiného, předložit veřejnosti ucelenou studii k daným otázkám v českém jazyce, neboť podobná publikace – ve srovnání se zahraničím, prozatím u nás chyběla. Tohoto úkolu se ujal s velkou odpovědností a pílí sobě vlastní. Publikace má kompaktní ráz, fundovaný obsah, který je podán v precizní formě a vyznačuje se vysokou mírou logické uspořádanosti.

Obsah recenzované knihy je z hlediska metodologického postaven především na velmi pečlivém studiu a vlastní interpretaci a hodnocení podstatné zahraniční (v menší míře i domácí) literatury a dalších zdrojů vysoce relevantních sledovanému tématu. Závěrečný seznam použité literatury čítá přibližně 130 titulů. Pokud se podíváme na počet citací k jednotlivým zaznamenaným titulům, pak se jako nosné prameny, které autor použil pro formulaci textu recenzované knihy, jeví následující (dokládají zároveň, že oblast digitální archivace je prozatím doménou vybraných zahraničních zemí):

- zásadní dokument normativní povahy globálního významu „Referenční model otevřeného archivního informačního systému – OAIS“, který v roce 2002 na svém serveru volně zveřejnil Konzultační výbor pro systémy dat z oblasti vesmíru (<http://public.ccsds.org/publications/archive/650x0b1.pdf>) a který po roce vyšel také jako mezinárodní norma ISO 14721:2003 (autor recenzované knihy volně zveřejněnou verzi standardu citoval 30x)
- významná rozsáhlejší odborná monografie k problémům archivace a ochrany digitálních informací amerického počítačového experta Henryho M. Gladneye s anglickým názvem

„Preserving Digital Information“, publikovaná vydavatelstvím Springer v roce 2007 (je autorem citována 20x)

- přehledová publikace k otázkám digitální archivace a ochrany dokumentů z praxe vybraných evropských národních knihoven nizozemské expertky Ingeborg Verheulové s anglickým názvem „Networking for digital preservation : current practice in 15 National libraries“, vydaná vydavatelstvím Saur v roce 2006 a později zveřejněná volně také na serveru asociace IFLA (<http://archive.ifla.org/VI/7/pub/IFLAPublication-No119.pdf>) (je autorem citována 14x)
- užitečná příručka k otázkám managementu archivace digitálních materiálů dvojice známých britských autorů (Neil Beagrie a Maggie Jonesová), která byla s anglickým názvem „Preservation management of digital materials“ zveřejněna v roce 2008 na serveru koalice DPC (<http://www.dpconline.org/pages/handbook/index.html>) (je autorem citována 11x; URL uvedené autorem v knize již není funkční)
- standard datového modelu a specifikace archivačních metadat PREMIS ve verzi 2.0 z roku 2008, zveřejněný Kongresovou knihovnou (<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>) (je autorem citován 10x)
- významná zpráva (Preserving digital information) společné Pracovní skupiny pro archivaci digitálních informací Komise pro ochranu a přístup Rady pro knihovnictví a informační zdroje (CLIR) a bývalé sítě RLG (obě USA), vydaná v roce 1996 a zveřejněná volně také na serveru CLIR (USA) (<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>) (je autem citována 8x)

Autor z analyzovaných pramenů dokázal vybrat podstatné informace, které podle možnosti porovnává mezi sebou, vyhodnocuje a konfrontuje s dílčími zkušenostmi získanými v praxi mateřského pracoviště NK ČR. Jeho práce s terminologií oblasti digitální archivace je velmi promyšlená, ocenit je potřebné snahu o důslednost při překládání specifických termínů do českého jazyka, byť to bylo v několika případech objektivně složité, takže se autor uchýlil k pracovnímu zavedení některých anglicismů do české terminologie (například termínu „stakeholders“, „resolver“ aj.). Řada termínů je nově uvedena, jako termín „digitální elektronická informace“ aj. O takových termínech je možné dále debatovat, i proto autor knihu, jak lze předpokládat, předkládá veřejnosti k posouzení.

Hlavní text publikované práce je, kromě preliminárií a postliminárií, autorem uspořádán do 10 základních kapitol, přičemž poslední kapitola je věnována závěrečnému souhrnu sledované problematiky. Struktura celé práce je hluboce promyšlená a propracovaná, v některých kapitolách jde autor až na pátou úroveň dělení probírané látky. Titulky obsahu v zásadě vystihují dílčí témata otázek digitální archivace velmi přesně. V celku publikace lze rozeznat tři významné části. Prvních pět kapitol tvoří výchozí část celé práce, zahrnuje vymezení celé problematiky dlouhodobé archivace digitálních dokumentů a představuje sérii relevantních termínů. Hlavní jádro práce tvoří kapitoly šestá až osmá, autor v nich postupně představuje hrozby či překážky digitální archivace, vlastní interpretaci výchozích tezí referenčního modelu OAIS a zejména pak řadu existujících ochranných opatření digitální archivace. Kapitola 9. má charakter hodnotícího závěru celé problematiky.

V 1. kapitole, která je úvodem do tématu, autor stručně představuje otázky dlouhodobé ochrany dokumentů včetně dokumentů digitálních. Podtrhuje význam role elektronické informace a komunikace ve společnosti, zároveň ale správně a s důrazem poukazuje na nové hrozby této formy komunikace informací, jež jsou společností zatím jen velmi málo vnímány. Autor zajímavým způsobem konfrontuje specifika ochrany dokumentů tradičních a dokumentů digitálních a pojmenovává problémy z tohoto srovnání vyplývající. Naléhavou podmínkou trvalé udržitelnosti digitálních informací je připravit novou koncepci jejich ochrany. Autor v rámci 2. kapitoly, která podává genezi celého problému, vyzývá k přípravě a rozvíjení nových ochranných opatření. Pomoci by měl také nově se konstituující obor dlouhodobé ochrany digitálních dokumentů (digitální archivace). Tento obor autor podrobněji představuje v 3. kapitole. Pečlivě analyzuje a srovnává definice digitální archivace renomovaných zahraničních autorů (např. N. Beagrie a H. M. Gladney), podtrhuje úlohy tohoto oboru, popisuje jeho vznik a nosný pilíř, jímž je zpracování a zveřejnění mezinárodní normy ISO 14721:2003 (referenční model OAIS). V podkapitole 3.3 autor v kvalitně zpracovaném přehledu představuje významné výzkumné projekty v oblasti digitální archivace v globálním rámci včetně institucí, které je zaštiťují. V podkapitole 3.4 je autorem dokladována institucionalizace tohoto nového oboru (vznik specializovaných periodik, odborných institucí a organizací a konferencí).

Čtvrtá kapitola přináší v zajímavém uspořádání zasvěcený výklad základních termínů spojených s digitálním dědictvím. Autor poukazuje na nové paradigma jeho ochrany, zdůrazňuje nutnost spolupráce všech paměťových institucí, představuje obecné cíle jeho trvalé udržitelnosti a vymezuje vztahy k příbuzným oborům (digitalizace a digitální

knihovny). V navazující 5. kapitole předkládá detailnější výklad termínu „digitální dokument“ (v části 5.7 se nachází také jeho základní typologie) a termínů s ním souvisejících. Již v úvodní 1. kapitole autor použil specifický termín „digitální elektronická informace“ – předpokládám, že záměrně a výjimečně, v praxi není tato kombinace obou přídavných jmen, byť je logicky správná, běžně používaná a ani autor ji v dalším textu nepoužívá. Pravděpodobně v souvislosti s tím ale autor osvětluje sémantický význam obou přídavných jmen (v části 5.1). S výkladem lze zcela souhlasit. V dalších částech je vyloženo pojetí informace podle standardu OAIS, dále problematika granularity zpracovávaných, uchovávaných a zpřístupňovaných dokumentů, zejména digitálních, a konečně termín „digitální objekt“. V souvislosti s ním je charakterizován datový model metadatového standardu PREMIS (autor ho pokládá za optimální pro potřeby zpracování digitalizovaných dokumentů v knihovnách) a také technický model digitálního objektu (podle K. Thibodeaua).

Jak jsem uvedla výše, za hlavní jádro Cubrovy práce lze pokládat obsahy kapitol 6.-8. Jsou velmi přínosné a pro čtenáře i dobře čtivé (zaregistrují postupně sérii známých i méně známých či neznámých skutečností, často v nových souvislostech s hodnotícími stanovisky autora). Šestá a osmá kapitola spolu úzce souvisejí, a proto mají identickou strukturu uspořádání na druhé úrovni dělení. V rozsáhlejší 6. kapitole autor podává v souladu se světovým publikovaným poznáním kvalitní, dobře utříděný přehled problémových rizik (hrozeb) dlouhodobé ochrany digitálních dokumentů, resp. překážek jejich trvalé udržitelnosti. Předkládané informace jsou často porovnávány a hodnoceny s řadou doplňujících vlastních stanovisek autora. Vykládané otázky jsou vhodně vymezeny ve čtyřech rovinách problému: 1. technologické (jde zejména o vážnou hrozbu degradace nosičů a zastarávání hardwaru), 2. informační (ohroženy jsou datové formáty digitálních objektů v souvislosti s prudkým vývojem technologií a aplikačních programů), 3. systémové (hrozba nezvládnutí optimální správy enormního množství digitálních objektů, jejich přesné identifikace, zachování integrity a autenticity a také zajištění práv k jejich využívání) a 4. institucionální (hrozba ukládání digitálních dokumentů do nevyhovujících (nedůvěryhodných) digitálních archivů (autor užívá termínu „digitální repozitáře“), hrozba nedostatečného finančního a organizačního zajištění provozu těchto systémů aj.).

V 7. kapitole autor předkládá kvalitně připravený vlastní výklad základů známého referenčního modelu OAIS (ISO 14721:2003) doprovázený sérií schémat. Autor zdůrazňuje význam tohoto modelu pro budování systémů pro dlouhodobou archivaci digitálních dokumentů a blíže popisuje pojetí informace (informační model) a funkční model standardu

OAIS. Texty jsou vhodně doprovázeny obrázky (s českými texty). K této části bych měla následující poznámky. Autor vykládá modely v zásadě jenom z pohledu digitálního prostředí (kvůli tomu také mění některé termíny, například namísto termínu „OAIS (archiv)“ užívá termín „repozitář“, namísto termínu „spotřebitel“ užívá termín „klient“ aj.), zatímco citovaný standard OAIS je veden naprosto v obecné rovině – postihuje jakékoliv systémy, tradiční a digitální (jde o zásadní princip přijatý jeho tvůrci). Proto by bylo vhodné, kdyby autor promítl tuto skutečnost do názvu 7. kapitoly a do poznámek pod čarou, jinak nezasvěcený čtenář může nabýt dojmu, že to tak je v originálním dokumentu. Za další, autor přebírá vybraná schémata z originálního dokumentu (některá možná i vytvořil zcela nově, nebo jde o výřezy z širších originálních schémat), cituje je však neúplně (chybí buď označení a číslo obrázku, nebo alespoň strana). Například autorovo schéma 2 koresponduje s obrázkem originálu (Figure) 2-1, je však modifikováno terminologicky. S tím souvisí poslední připomínka – v legendách ke schématům by bylo vhodné přesně uvést, že jde o „úpravu“ či „modifikaci“ podle originálu.

Významnou a přínosnou částí práce je též 8. kapitola. V ní autor ve velkých detailech a se zaujetím představuje základní praktická ochranná opatření dlouhodobé archivace digitálních dokumentů. Struktura dílčích podkapitol je, jak bylo uvedeno výše, identická se 6. kapitolou, ochranná opatření jsou sledována opět ve čtyřech rovinách. Výklady jsou nabitě mnoha fakty, autor vychází z publikované literatury, ze zveřejněných informací z různých projektů prezentovaných na WWW a také zkušeností z vlastního pracoviště NK ČR. Texty jsou analytické a velmi podrobné. Obsahy jsou poučné.

Předposlední 9. kapitola je podle mého názoru precizně formulovaným závěrem celé problematiky dlouhodobé digitální archivace. Autor konstatuje, že rizika technologické roviny dnes nejsou již hlavním problémem digitální archivace a že k základním ochranným opatřením patří zejména soustavná bitová ochrana. Formátová rizika jsou závažnějším problémem, autor interpretuje složitý spor odborníků z 90. let k ochranným opatřením formátové migrace a emulace a ztotožňuje se s tezí, že optimálním východiskem je dnes prosazování komplexu různých ochranných opatření dohromady. Autor také vítá nové formátové strategie, ke kterým patří online formátové registry, a vývoj nových formátů pro digitální archivaci s důrazem na formáty, které mají otevřenou specifikaci. Poukazuje na důležitou roli metadatových specifikací, zejména pak archivačního, technického, právního charakteru. Zdůrazňuje trend posunu od jednoduchých ke komplexním ochranným opatřením digitální archivace. Publikace je ukončena v 10. kapitole vhodným hodnotícím souhrnem celé

problematiky. Autor apeluje na rozumné řešení překotného procesu digitalizace informačních fondů v situaci, kdy společnost nemá vyjasněné základní otázky dlouhodobé archivace digitálního dědictví a jeho trvalého zpřístupňování. S autorem lze v tomto směru zcela souhlasit.

Zveřejněná publikace Ladislava Cubra má vedle nesporných obsahových kvalit také (dovolují si uvést mimořádné) kvality formální (měly je již texty diplomové a rigorózní práce). Není často vidět v textech absolventů oboru tak skvělá čeština a stylistika. Vynikající úroveň má poznámkový aparát pod čarou. Závěrečná redakce lektorů a redaktorů neodhalila jen čtyři drobné chyby (uvádím je pro potřeby případných dalších vydání textu): na s. 144 v rozpisu jména firmy IBM je chyba ve slově „Machine“, správně „Machines“ a poznamenávám, že forma korporativního záhlaví ale není v souladu s citací na s. 97 „[IBM, 2009]“; na s. 145 forma korporativního záhlaví není v souladu s citací na s. 132 „[IFLA, 2009]“; na s. 148 v rámci záznamů začínajících slovem „nestor“ je u druhého záznamu v začátku trvalého identifikátoru nesprávné označení „um“ místo správného „urn“ a na s. 152 v rozpisu zkratky CCSDS je chyba ve slově „Commitee“, správně „Committee“.

Publikovaná monografie Ladislava Cubra je podstatným souhrnným příspěvkem k problematice trvalé udržitelnosti digitálního dědictví v českém jazyce. Autor v jádru práce vytěžil podstatné informace ze světové literatury, informace zajímavě uspořádal a představil v kvalitně napsaném textu. Publikace by neměla chybět v odborných příručkách profesionálů našeho oboru, je vhodným textem i pro potřeby relevantní vysokoškolské výuky (v tomto směru již slouží i pro výuku na ÚISK) a školení zájemců o tuto problematiku.