

Cubr, Ladislav

## **Konceptuální rámec pro otázky autenticity digitalizátů knih**

*ProInflow*. 2018, vol. 10, iss. 2, pp. 70-107

ISSN 1804-2406 (online)

Stable URL (DOI): <https://doi.org/10.5817/ProIn2018-2-5>

Stable URL (handle): <https://hdl.handle.net/11222.digilib/138450>

License: [CC BY 3.0 CZ](#)

Access Date: 16. 02. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

# KONCEPTUÁLNÍ RÁMEC PRO OTÁZKY AUTENTICITY DIGITALIZÁTŮ KNIH

## CONCEPTUAL FRAMEWORK FOR THE AUTHENTICITY OF DIGITIZED BOOKS

Ladislav Cubr

*Národní knihovna České republiky*

*Realizováno v rámci institucionálního výzkumu Národní knihovny České republiky financovaného Ministerstvem kultury ČR  
v rámci Dlouhodobého koncepčního rozvoje výzkumné organizace.*

### Abstrakt

**Účel** – Konceptní článek se zabývá fenoménem autenticity digitálních dokumentů. Podrobněji se zaměřuje na otázku autenticity digitalizátů knih jakožto specifického typu dokumentu, která není v současné odborné literatuře soustavně konceptuálně uchopena. Cíle článku jsou následující. Prvním cílem je zmapovat šíři otázek souvisejících s konceptem autenticity ve vztahu k řízení životního cyklu digitalizátů knih v rámci jedné organizace, a to na základě revize existujících vybraných dílčích tematizací pojmu autenticita v oblasti paměťových institucích. Druhým a hlavním cílem je navrhnout základní konceptuální rámec pro postupy udržování a posuzování autenticity digitalizátů knih.

**Design/metodologie/přístup** – V úvodu článku představujeme pojetí MacNeilové a Makové, ve kterém je autenticita sociální konstrukt – v závislosti na užití perspektivě vedle sebe mohou legitimně existovat odlišné konceptualizace autenticity téhož předmětu. Následně přibližujeme pojetí konceptu autenticity digitálního dokumentu v archivnictví a základní rámec pro životní cyklus digitálních dokumentů obsažený v modelu OAIS. Tyto dva zdroje jsou užity jako výchozí rámec pro následnou kritickou a konceptuální analýzu odborných zdrojů. Cílem této analýzy je určit základní konceptuální prvky, které se vztahují k pojmu autenticita v kontextu celého životního cyklu digitalizátů knih. Výstupy této analýzy jsou pak základem pro návrh konceptuálního rámce pro postupy udržování a posuzování autenticity digitalizátů knih. Tento návrh je založen na premise, že pro konkrétní typ dokumentu je nutno vypracovat specifickou konceptualizaci autenticity.

**Výsledky** – Hlavním výsledkem předkládaného článku je návrh základního konceptuálního rámce pro autenticitu digitalizátů knih. Tento rámec zahrnuje redefinici pojmu autenticity digitálního dokumentu, zavedení nového pojmu „referenční objekt“ a návrh specifikace minimálních atributů nezbytných pro vyjádření identity digitalizátu knihy. Tyto koncepty jsou základním předpokladem pro procesy udržování a posuzování autenticity digitalizované knihy. Konceptuální rámec ukazuje na možnost několika odlišných přístupů k autenticitě digitalizátů knih. Tím se potvrzuje koncept MacNeilové a Makové o legitimní koexistenci odlišných konceptualizací autenticity téhož předmětu.

**Originalita/hodnota** – Dosud neexistuje soustavná konceptualizace autenticity digitalizátů knih, která by zohledňovala všechny aspekty a fáze životního cyklu dokumentu. Tento příspěvek představuje autorův vlastní návrh základního konceptuálního rámce pro otázky udržování a posuzování autenticity digitalizátů knih, který je ve své komplexitě jedinečný.

**Klíčová slova** – autenticita, identifikace, digitální dokument, digitalizát knihy, digitalizace, archivace, OAIS

**Abstract**

**Purpose** – The article addresses the phenomenon of authenticity of digital documents and focuses in more detail on the issue of the authenticity of digitized books as a specific type of document. Currently there is no systematic conceptualization of the authenticity of digitized books in research literature. There are two objectives of the article. The first objective is to map the scope of the issues pertaining to authenticity in the context of the life cycle management of digitized books within an organization. For this purpose, the selected literature from the field of memory institutions is reviewed. The second and the main objective is to suggest an underlying conceptual framework for maintaining and assessing the authenticity of digitized books.

**Design/methodology/approach** – The introduction of the article presents the concept of authenticity as a social construct developed by MacNeil and Mak. According to this concept, different authenticities of the same object may legitimately coexist depending on the perspective used. Then the archival concept of the authenticity of the digital document and the OAIS high-level framework for the digital documents' life cycle are introduced. Both resources are used as a basis for subsequent critical and conceptual analysis of literature. The analysis seeks for identifying fundamental conceptual elements pertaining to the concept of authenticity in the context of the life cycle of the digitized books. The outcomes of the analysis are the starting point for suggesting a conceptual framework for maintaining and assessing the authenticity of digitized books. This framework is based on the premise that establishing specific concept of authenticity is required for each type of document.

**Results** – The main outcome of the paper is an underlying conceptual framework for the authenticity of digitized books. The suggested framework includes a redefinition of the concept of authenticity of the digital document, the introduction of the new term “reference object”, and the specification of minimal attributes needed for expressing digitized book identity. All these components are the fundamental prerequisite for maintaining and assessing the authenticity of the digitized books. The conceptual framework implies that various approaches to the authenticity may be employed. This confirms the McNeal's and Mak's concept of legitimate co-existence of different authenticities of the same subject.

**Originality/value** – There is still no systematic conceptualization of the authenticity of the digitized books describing all aspects and phases of the life cycle of the document in research literature. This paper presents the author's draft of an underlying conceptual framework for maintaining and assessing the authenticity of digitized books, which is unique in its complexity.

**Keywords** – authenticity, identification, digital document, digitized book, digitization, archiving, OAIS

## 1 Úvod

S konceptem autenticity se setkáme v řadě nejrůznějších disciplín. Zatímco digitální informace se staly celospolečenským fenoménem koncem dvacátého století, pojem autenticity a její sociokulturní význam má mnohem hlubší kořeny. Autenticita v kontextu tradičních (fyzických) objektů byla (a je) klíčovým pojmem ve výtvarném umění, literární teorii, archivnictví, diplomacie a řadě dalších oblastí. V digitálním světě však platí, že vzhledem k nestabilní povaze digitálních objektů se autenticita stala kruciólním konceptem ve všech oblastech jejich užití bez ohledu na obor. Požadavek zaručit autenticitu digitálních dokumentů je zmiňován jako hlavní cíl digitální archivace (dlouhodobého uchovávání digitálních dokumentů navzdory změnám technologií) v celé řadě studií (Giaretta, 2011, s. 203). Výzkum autenticity v digitálním prostředí je však stále spíše ve svých počátcích. V oblasti digitálního dědictví existuje de facto jediná významná konceptualizace autenticity, a to v oblasti archivnictví.

Otázka autenticity fyzických dokumentů uchovávaných knihovnami nepředstavovala vážnější problém vzhledem k technologické obtížnosti jejich zfalšování a osvědčeným postupům získávání dokumentů z ověřených zdrojů (Deegan & Tanner, 2006; Bearman & Trant, 1998). Konceptualizace autenticity fyzických knihovních jednotek se zaměřují spíše na starší dokumenty, jako jsou rukopisy. Autenticita novodobých tištěných knih nebývá tematizována. Je tomu tak zřejmě vzhledem k malé motivaci zfalšování (řádově nižší cena knih ve srovnání s rukopisy), existenci více exemplářů téhož vydání a zavedenému institutu povinného výtisku (tj. získávání knih přímo od vydavatelů, a tedy z důvěryhodných zdrojů).

V průběhu devadesátých let minulého století se však otázka autenticity dokumentů dostala do popředí zájmu knihoven. Stalo se tomu tak v souvislosti s tím, jak tyto paměťové instituce začaly rozšiřovat své fondy o digitální objekty. Záhy se ukázalo, že jejich správa je mnohem problematičtější než u fyzických dokumentů. Digitalizáty knih se staly jedním z nejrozšířenějších typů dokumentů zastoupených v knihovních fondech a v praxi je to často jedna a tatáž organizace, která tyto dokumenty vytváří, uchovává i zpřístupňuje. Je tedy možné (a záhodné) na jejich správu aplikovat mechanismy soustavného řízení celého životního cyklu dokumentu (od plánování digitalizačního projektu přes produkci a archivaci až k trvalému zpřístupňování). Dosud však v knihovnách nevznikla soustavná konceptualizace autenticity digitalizátů knih, která by popisovala postupy pro udržování a posuzování jejich autenticity v rámci celého tohoto řízení.

Účelem tohoto článku je navrhnout základní konceptuální rámec pro autenticitu digitalizátů knih. Vycházíme z premisy, že pro konkrétní typ dokumentu je nutno vypracovat specifickou konceptualizaci autenticity, kterou nalezneme u několika autorů (např. Bearman & Trant, 1998; Gladney, 2007). Úvodním krokem pro návrh konceptuálního rámce je představení modelu pro dlouhodobý životní cyklus dokumentů (model OAIS) a tradiční definice autenticity z oblasti archivnictví, které se staly výchozím bodem pro celkový rámec; druhým krokem pak revize vybrané odborné literatury s cílem identifikovat celou šíři možných konceptuálních prvků, které jsou relevantní pro otázky autenticity digitalizátů knih. Na této analýze je pak založen navrhovaný konceptuální rámec.

## 2 Výchozí definice autenticity a rámec OAIS

Za vhodný úvodní vstup do obecné problematiky autenticity obecně považujeme koncepci autenticity jako sociální konstrukt, kterou představily MacNeilová a Maková v článku v *Library Trends* (2007). Autorky vycházejí z definice termínu „authentic“ v *Oxford English Dictionary* (2018), který uvádí tři významy: „pravdivý sám k sobě“ (*true to oneself*), „původní“ (*original*) a „důvěryhodný“ (*trustworthy*) ve vztahu k tvrzení o skutečnosti. Podobné definice nalezneme v českých zdrojích. Například *Nový akademický slovník cizích slov* definuje pojem „autenticita“ jako „pravost, původnost, hodnověrnost“; adjektivum „autentický“ analogicky jako „původní, pravý, hodnověrný“ (Kraus & et al., 2005). Totožné definice nalezneme v řadě dalších slovníků a encyklopedií [např. (*Malá československá encyklopedie*, 1984-1987; Kroupová & et al., 2005; Vácha & et al., 1971)].

MacNeilová a Maková rozebírají uvedené kategorie významu pojmu autenticita ve vybraných teoretických pracích z oblasti filozofie, výtvarného umění a literární teorie a teorie práva. Ve všech z nich je podle autorek autenticita (*authenticity*) chápána jako cíl (kýžená vlastnost) stanovený pro určitý předmět, který má být zachován navzdory specifickým rizikům nebo překážkám. Ve filozofii jde o zachování si jedinečnosti individuality osoby navzdory společenskému tlaku na uniformitu; v oblasti výtvarného umění o zachování / rekonstruování původního díla navzdory jeho degradaci a v kontextu otázek, jak dílo adekvátně restaurovat; v literární teorii / teorii textu o rekonstruování původního (již zaniklého) textu z jeho pozdějších (dochovaných) prepisů nebo hledání autorského záměru navzdory vydavatelským úpravám; v teorii práva o prověřování a posuzování důkazů u soudu v kontextu odlišných nebo neúplných svědectví. Autorky dospívají k velmi obecnému závěru: „autenticita je sociální konstrukt, který je využíván v řadě disciplín k tomu, aby pomáhal strukturovat jejich konkrétní prostředí“ (MacNeil & Mak, 2007, s. 44). Podle nich mohou vedle sebe legitimně existovat odlišné konceptualizace autenticity jednoho a téhož předmětu, a to v závislosti na tom, jaká perspektiva nebo teorie byla zvolena.

Obecné kategorie významu (pravdivý / pravý; původní; hodnověrný / důvěryhodný) mohou posloužit jako vstupní orientační rámec pro problematiku autenticity. Pro konceptualizaci autenticity digitalizátů knih v kontextu jejich správy v knihovnách je však nejprve potřeba zúžit oblast zkoumání na otázku autenticity specifického objektu, jímž je dokument. Dokument je v tomto článku chápán v obecnějším významu jako jakýkoliv institucionalizovaně spravovaný datový objekt (knih, mapa, socha, obraz, filmový pás apod.), který je evidován a uchováván určitou organizací pro informační potřeby nějaké skupiny uživatelů. V problematice autenticity dokumentu lze pak za výchozí bod považovat pojetí tohoto pojmu v archivnictví. Právě to se stalo základem pro užívání pojmu autenticita v knihovnách a dalších paměťových i výzkumných institucích.

Podle Luciany Durantiové, kanadské archivní teoretičky, sahá koncept autenticity dokumentu až do starořímské právní teorie, odkud byl v 17. století převzat do klasické diplomatiky a později do archivní vědy (2001). Durantiová působila jako vedoucí výzkumník v prvních projektech věnovaných konceptualizaci autenticity digitálních dokumentů: *Preservation of the Integrity of Electronic Records* kanadsko-

americké provenience a v mezinárodním *International Research on Permanent Authentic Records in Electronic Systems (InterPARES)*. Tyto projekty se specificky zaměřovaly na určitý typ dokumentu (archiválii).

První projekt zkoumal integritu (*integrity*) dokumentů, které jsou využívány v běžné činnosti původce. Integritu pojal jako celek dvou složek: spolehlivosti (*reliability*) a autenticity. Spolehlivý dokument je v projektu definován jako takový dokument, se kterým „lze zacházet jako s faktem samotným, to jest jako s entitou, jejímž je dokladem (*evidence*).“ (Duranti, 1995, s. 6); spolehlivost dokumentu znamená jeho důvěryhodnost (*trustworthiness*) jakožto dokladu, tj. „schopnost vypovídat o skutečnostech, o kterých pojednává.“ (Duranti, 1996, s. 54). Dokument je pak tehdy „autentickým, pokud je takovým dokumentem, za který se prohlašuje (*it claims to be*).“ (Duranti, 1995, str. 7). Spolehlivost dokumentu a způsoby jejího zaručení jsou podle projektu spojeny výhradně s procesem vytváření dokumentu (otázky míry úplnosti dokumentu a řízení postupu jeho vytváření). Autenticita dokumentu znamená, že dokument je právě tak spolehlivý, jako byl spolehlivý v době svého vzniku (i nespolehlivý dokument tedy může být autentický): udržení autenticity podle projektu „pouze zaručuje, že dokument není výsledkem manipulace, nahrazení nebo zfalšování, ke kterému by mohlo dojít po dokončení procesu jeho vytvoření, a tedy že [dokument] je tím, za co se vydává (*what it purports to be*).“ (Duranti, 1995, s. 7). Hlavní úkol původce spočívá v dodržování takových postupů, které zaručí vytváření spolehlivých digitálních dokumentů v organizaci: „místo toho, abychom se soustřeďovali na zajištění integrity nespolehlivých dokumentů, měli bychom se zaměřit na zajištění vytváření spolehlivých dokumentů“ (Duranti, 1995, s. 9). Pro zkoumání otázek autenticity vyšel projekt z následující logiky: aby bylo možno stanovit požadavky pro zachování a posuzování autenticity digitální archiválie, je nejprve potřeba vytvořit model tohoto typu dokumentu. Za tímto účelem byly převzaty koncepty klasické diplomatiky a vytvořen univerzální model archiválie. Hlavní výstupy projektu lze shrnout do následujících závěrů:

- a) tentýž model dokumentu lze užít jak pro tradiční, tak digitální dokumenty, s tím rozdílem, že konstituenty digitálního dokumentu existují odděleně – udržení autenticity digitálního dokumentu tedy vyžaduje jednotně řízenou správu všech jeho konstituentů;
- b) základní přístupy tradičního uchovávání (kopírování, reformátování) se uplatňují i v digitální archivaci;
- c) tradiční princip, podle kterého spoléhání se organizace při výkonu své činnosti na kopie dokumentů činí tyto kopie autentickými, je platný i pro digitální dokumenty.

Navazující projekt InterPARES se zaměřoval na uchovávání archiválií ve specializovaných organizacích (tedy vybraných dokumentů, které původce přestal užívat a které byly vybrány k dlouhodobé archivaci). Projekt vytvořil nový model dokumentu, jehož třemi konstituenty jsou dokumentová forma, anotace a kontext. V otázce způsobu uchovávání dospěl InterPARES k zcela opačnému závěru – uchovávání digitálních dokumentů se od archivace fyzických dokumentů paradigmaticky odlišuje: „uchovávat digitální dokument není možné: lze pouze uchovávat schopnost dokument reprodukovat (*reproduce*)“ (InterPARES project, 2005b). Zatímco předchozí projekt pojímal autenticitu jako jednu ze složek integrity, v tomto

projektu je autenticita definována jako celek identity (*identity*) a integrity (*integrity*). Identita dokumentu jsou „atributy dokumentu, které jej jedinečně charakterizují a odlišují od jiných dokumentů“ (InterPARES project, 2005c, s. 1); integrita dokumentu se týká jeho úplnosti (*wholeness*) a neporušenosti (*soundness*): „dokument má integritu, pokud je úplný (*complete*) a neporušený (*uncorrupted*) ve všech podstatných aspektech (*essential aspects*).“ (InterPARES project, 2005c, s. 2). To podle projektu znamená, že nesmí být pozměněno sdělení, které má dokument přenášet s cílem dosáhnout určitého účelu a zároveň musí být zachováno vyjádření obsahu a všechny požadované anotace a znaky dokumentové formy (InterPARES project, 2005c, s. 2). Spojení „podstatné aspekty“ tedy znamená připuštění určitých změn digitálního dokumentu, které nenarušují autenticitu. Posouzení autenticity znamená „určit (*establish*) jeho identitu a prokázat (*demonstrate*) jeho integritu“ (InterPARES project, 2005c, s. 1), a to na základě dodatečných dokladů o způsobu vytváření a udržování dokumentů původcem (InterPARES project, 2005c, s. 3). Pro (následně) udržení autenticity dokumentu v archivu stanovil projekt sadu požadavků, z nichž za hlavní lze označit zachování podstatných aspektů po skončení procesu vytváření kopií a dokumentaci tohoto procesu včetně jeho důsledků (InterPARES project, 2005c, s. 11).

Referenční rámec OAIS, obsažený v normě ISO 14721 (ISO, 2012a), se stal výchozím modelem pro většinu organizací usilujících o dlouhodobé uchovávání digitálních dokumentů ve svých repozitářích. Klíčovým pojmem normy je „informační obsah“ (*content information*), což je objekt, který je hlavním cílem dlouhodobého uchovávání; tento pojem zhruba odpovídá konceptu dokumentu v archivnictví. Informační obsah se skládá z objektu s datovým obsahem (*content data object*; dále jako „objekt CDO“) a interpretačních informací (*representation information*) (ISO, 2012a). Podle normy musí být informační obsah uchováván v repozitáři společně s dodatečnými informacemi, které jsou označeny jako archivační informace (*preservation description information*). Příkladem objektu CDO může být soubor ve formátu PDF představující knihu. Interpretační informací bude například formátová specifikace, podle níž lze vytvořit software k zobrazení textu e-knihy (tedy informačního obsahu) zakódovaného v PDF. Pokud vztáhneme tento model na běžnou digitalizaci, pak objekt CDO je často tvořen třemi komponenty: obrazovou (soubory ve formátech TIFF nebo JP2, z nichž každý obvykle odpovídá jedné straně knihy), textovou (soubory ve formátu ALTO, z nichž každý odpovídá jedné straně knihy a které jsou výstupem OCR a zajišťují plnotextovou prohledatelnost) a strukturální (strukturální mapa, která je obsažena v části schématu METS a která vyjadřuje vztahy mezi komponentami a jejich částmi, jednotlivými soubory). Interpretační informace jsou pak informace, potřebné ke správnému zobrazení digitalizátů knihy (např. popis formátů a nástrojů pro jejich zobrazení).

Co je překvapující, ale do značné míry příznačné, je skutečnost, že v prvním vydání normy (ISO, 2003) není pojem autenticita vůbec definován. To byl také jeden z hlavních nedostatků normy, jak později přiznal sám David Giarretta, hlavní autor (2011, s. 205). Definici autenticity zavádí až druhé vydání (ISO, 2012a), přičemž vliv zmíněných archivářských projektů je více než zjevný. V terminologickém slovníku normy se k pojmu autenticita uvádí: „Míra, do které osoba (nebo systém) pokládá objekt za takový, za který je vydáván (*The degree to which a person (or system) regards an object as what it is purported to be.*); autenticita se

posuzuje podle dokladů (evidence)“ (ISO, 2012a, s. 20). Úkol zachovat autenticitu je v nové verzi normy zahrnut i do samotného pojmu „dlouhodobé uchovávání“ (*long term preservation*), které je nyní definováno jako: „dlouhodobé udržování informací v podobě, která je cílové komunitě srozumitelná sama o sobě, spolu s doklady o jejich autenticitě“ (ISO, 2012a, s. 24). Požadavky na udržení autenticity norma vztahuje na všechny typy informačních objektů uchovávaných v repozitáři, ale vzhledem k ústřední pozici informačního obsahu lze logicky odvodit, že jádrem všech procesů pro udržování autenticity je zachování autenticity informačního obsahu jakožto nejdůležitějšího typu informačního objektu. Hlavním zdrojem dokladů o autenticitě informačního obsahu jsou provenienční informace (*provenance information*): „informace, které dokumentují historii informačního obsahu. Tyto informace popisují původ nebo zdroj informačního obsahu, všechny změny, které by provedeny od jeho vzniku, a kdo byl správcem informačního obsahu“ (ISO, 2012a, s. 25). Koncept provenienčních informací<sup>1</sup> a definici autenticity tak norma ISO 14721:2012 zjevně z větší části přebírá z výstupů projektu InterPARES.

### 3 Konceptuální prvky autenticity

V této práci byl jako výchozí konceptuální rámec pro popis životního cyklu dokumentu pro navrhovanou konceptualizaci autenticity zvolen referenční rámec OAIS, a to vzhledem k dostatečné obecnosti a výstižnosti a celosvětovému rozšíření.

Ačkoliv ve zdrojích z oblasti knihovnictví nenalezneme žádnou konceptualizaci autenticity digitalizovaných knih, existuje zde celá řada jejích tematizací.<sup>2</sup> Za účelem zmapování šíře a specifik problematiky autenticity v kontextu řízení životního cyklu digitalizátů knih byly analyzováno 8 směrnic (Caplan, 2008; Data Seal of Approval Board, 2016; Digital Library Federation, 2002; Digital Preservation Handbook, 2018; Federal Agencies Digitization Initiative, 2010; Fernie, 2008; NISO, 2007; PREMIS, 2015; UNESCO, 2016) a 8 odborných publikací (Bearman, 1999; Bearman & Trant, 1998; Deegan & Tanner, 2006; Gladney, 2007; Grycz, 2006; Rothenberg, 1999; Rothenberg, 2000), které se věnují digitalizaci, archivaci nebo zpřístupnění digitálních dokumentů, resp. příbuzným oblastem, a přímo nebo nepřímo tematizují také některý aspekt autenticity.

Zkoumané zdroje se zabývají vždy alespoň jedním ze tří základních okruhů: a) definice autenticity (podstata autenticity, její předmět, složky apod.); b) udržení autenticity (předpoklady, podmínky, omezení, postupy apod. pro udržení / zachování autenticity objektu v jeho životním cyklu); c) ověření / posouzení autenticity (jakým způsobem / do jaké míry lze ověřit / posoudit, že datový objekt je autentický).

Jen sedm zdrojů uvádí definici autenticity (Caplan, 2008; Bearman & Trant, 1998; Digital Preservation Handbook, 2018; Gladney, 2007; NISO, 2007; PREMIS, 2015; UNESCO, 2016). Tyto zdroje - s výjimkou Gladneyho a směrnice NISO – však pouze parafrázuji tradiční definici autenticity v archivnictví

---

<sup>1</sup> V modelu dokumentu InterPARES je jedním z typů kontextu archiválie provenienční kontext.

<sup>2</sup> Tematizací zde rozumíme jakékoli pojednání, které není soustavné (tj. dílčí uchopení problému jen z některého aspektu, např. autenticity z hlediska věrnosti digitalizace).



(autenticita jako vlastnost dokumentu, který je tím, za co se vydává). Podle směrnice NISO je autenticita „stupeň důvěry, kterou může mít uživatel v integritu a důvěryhodnost objektu“ (NISO, 2007, s. 55). Pojmy integrita a důvěryhodnost však směrnice nedefinuje. Gladneyho definice je zdaleka nepracovitější. Tvrzení, že nějaký objekt je autentický (*authentic*), zahrnuje (2007, s. 97):

- a) tvrzení o odvození (*derivation statement*): „Objekt V je kopií objektu Y.“
- b) tvrzení o provenienci (*provenance statement*): „Tvůrce X vytvořil objekt Y jako součást události Z.“

Podle Gladneyho lze říci, že „objekt V je autentickou kopií objektu Y, pokud je dostatečně věrnou odvozeninou s pravou proveniencí.“ (2007, s. 97). Pravá provenience znamená, že tvrzení o odvození a tvrzení o provenienci jsou pravdivá; dostatečná věrnost znamená, že toto odvození odpovídá sociálním konvencím pro daný typ objektu a daný kontext. Gladney shrnuje: „Nejsme ochotni přijmout, že objekt V je autentický, pokud není explicitně určena minulé existence objektu Y, z něhož byl vytvořen objekt V, identita tvůrce objektu Y a konkrétní historická událost Z, v rámci níž k tomuto vytvoření došlo.“ (2007, s. 97).

Mnohem častěji zdroje uvádějí negativní vymezení autenticity, a to ve smyslu výčtu rizik pro její zachování. Zmíněná rizika lze shrnout do následujících kategorií: chyby při kopírování na jiný nosič (vč. přenosu po síti); nezáměrná chyba; záměna dokumentu za jiný; zfalšování; nedostatečná nebo chybějící dokumentace změn; formátová konverze (otázka emulace vs. migrace); chybná nebo nedostatečná identifikace; nejasný původ dokumentu; nekvalitní digitalizace.

Na základě analýzy uvedené literatury zdrojů bylo vyhodnoceno celkem deset základních konceptuálních prvků spojených s autenticitou: časoprostorový rámec; originál; aspekt odvozeniny; věrnost digitalizace; klíčové vlastnosti; identita; integrita; provenience; dokumentace; reprodukce počítačovou technologií.

Následující oddíly přinášejí revizi těchto kategorií, s případným vztažením k modelům OAIS a InterPARES a rozvedením specifických implikací pro digitalizáty knih.

### 3.1 Časoprostorový rámec

Rothenberg ve své studii uvádí: „V kontextu uchování (preservation) se autenticita inherentně vztahuje k času.“ (Rothenberg, 2000, s. 58). Podle Gladneyho pak „jakékoliv tvrzení o autenticitě srovnává nějakou entitu zde a nyní s něčím, co existovalo předtím.“ (Gladney, 2007, s. 97). Časový vztah je, přinejmenším implicitně, nutně přítomný ve všech tematizacích autenticity. Je zjevné, že musí existovat určitá časová distance mezi dvěma objekty (resp. stavy téhož objektu v různých časech), aby měl koncept autenticity smysl. Ve světě fyzických objektů je autenticita vztahem mezi aktuálním stavem fyzického objektu a jeho dřívějším stavem. V digitálním světě je důležitý nejen časový, ale i prostorový aspekt. Výstižné shrnutí časoprostorového rámce uvádí projekt InterPARES: autenticita digitálních dokumentů je „ohrožena, kdykoliv jsou přenášeny prostorem (tj. odesílány adresátovi nebo mezi systémy a aplikacemi) nebo časem (tj. během uložení v úložišti nebo aktualizace či náhrady hardwaru nebo softwaru užívaného k jejich

ukládání, zpracování nebo komunikaci).“ (InterPARES project, 2005c, s. 2). Vhodné je si také připomenout, že jakékoliv zpřístupnění digitálního dokumentu na internetu uživateli znamená vytvoření nové kopie dokumentu (zpřístupnění zakládá typ prostorového vztahu mezi objektem uloženým na serveru a jeho kopií staženou do zařízení uživatele).

Pro vyjádření všech typů časoprostorových vztahů nalezneme nejpropracovanější rámec v modelu OAIS, a to v konceptu informačních balíčků, při jejichž vytváření musí být podle normy zachována předepsaná posloupnost odvozování: SIP → AIP → DIP (ISO, 2012a). Pro procesy uchovávání lze tento vztah specifikovat jako vztah mezi aktuální (tj. poslední) a první verzí balíčku AIP (v různých verzích balíčku AIP se může objekt CDO lišit, typicky z hlediska jiného formátu) a dále také jako vztah mezi aktuální verzí balíčku AIP a jeho zálohami na dalších nosičích. Pro procesy zpřístupňování pak lze tento vztah specifikovat jako vztah mezi aktuální verzí balíčku AIP a balíčkem DIP (balíček DIP by, jak lze logicky odvodit z ustanovení normy, neměl být vytvářen ze starší verze balíčku AIP). Je třeba dodat, že rámec OAIS zahrnuje fázi uchovávání a zpřístupňování, nikoliv však počáteční fázi produkce.

### 3.2 Originál

Otázka originálu je v prostředí fyzických objektů s konceptem autenticity spojována pravděpodobně nejčastěji (Rothenberg, 2000). S konceptem originálu se pojí zejména rizika zfalšování nebo záměny, ale také surogátu dokumentu. Knihovny musejí již dlouho řešit otázku, jak uchovat knihy a periodika ohrožené degradací kyselého papíru. Deeganová a Tanner (2006) zmiňují případy pobouření britské veřejnosti nad opatřením knihoven odstranit rozpadající se fyzické originály a nahradit je mikrofilmy. Podle autorů je zničení originálu a jeho náhrada surogátem společensky citlivé téma, ale toto patření je pro knihovny ekonomicky i prakticky (z důvodu degradace) nevyhnutelné. Zdůrazňují, že vytvořením surogátu nelze nikdy zachovat všechno, co měl originál, ale nevytvořit žádný surogát může znamenat, že všechno bude ztraceno. Jinými slovy, fyzický surogát může zachovat alespoň část informačního obsahu originálu (text v jeho struktuře a vizuální vlastnosti zachycené na mikrofilmu).

Dále musíme odlišovat dva typy fyzických originálů: originál jako jedinečný objekt (např. Michelangelova socha Davida) a originál ve smyslu sériově vyráběného objektu (např. keramika). Německý filozof Walter Benjamin se ve slavné práci *Umělecké dílo ve věku své technické reprodukovatelnosti* (*Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit*) z roku 1936 zabývá kulturním významem prvně uvedeného typu originálu (Benjamin, 1979, s. 17-47). Zde představuje koncept jedinečného originálu (uměleckého výtvarného díla), který je časoprostorově situován do konkrétního (jedinečného) historického kontextu, ve kterém vznikl a jehož je odrazem. V tom také podle Benjamina spočívá jeho hodnota („aura“): toto „zde a nyní“ je vlastní pouze originálu a „vytváří pojem jeho pravosti (*Echtheit*). Chemické analýzy patiny na bronzu mohou napomoci ke stanovení pravosti; podobně jako k tomuto může vést důkaz, že určitý středověký rukopis pochází z archívu patnáctého století.“ (Benjamin, 1979, s. 3). Tento originál je jedinečný, a tedy nereprodukovatelný (zde myšleno ve smyslu, že nelze vytvořit

fyzickou kopii bez ztráty hodnoty). Tištěné knihy představují druhý typ originálu – „sériový“ originál. Všechny jednotlivé exempláře vydání knihy se považují za originál (tj. všechny exempláře jsou rovnocenné fyzické objekty). Ačkoliv kniha není jedinečný originál, tato skutečnost nikterak nesnižuje její hodnotu.

Bez ohledu na typ originálu dále platí, že fyzický originál, který je v současnosti k dispozici uživatelům, je vzhledem k nevyhnutelnému přírodnímu procesu degradace nutně jiný, než byl v době svého vzniku. Podle Gladneyho lze z tohoto hlediska označit aktuální stav fyzického objektu za „kopii“ objektu, kterým byl předtím (2007, s. 101), tj. kopii, která ztratila řadu původních znaků (např. původní bělost papíru knihy). Původní stav objektu lze pouze odhadovat na základě určité analýzy aktuálního objektu a jeho historie, přičemž tento dřívější stav může, ale nemusí být znám běžným percipientům daného objektu, kteří nejsou experti v dané oblasti.

Specifickým problémem je však i samo časové ukotvení originálu (ve smyslu původního stavu objektu); jak výstižně ukazují MacNeilová a Maková (2007, str. 31) na příkladu Michelangelova Posledního soudu, jeho restaurátoři museli řešit otázku, zda je původním stavem (a tedy předmětem restaurování) první verze díla z roku 1541, nebo pozdější autorovy úpravy (zakrývající nahé postavy, jak požadovala náboženská cenzura). Tuto distinkci můžeme nalézt, ač v méně závažné míře, i u tištěných knih. Exemplář v knihovně obsahuje prvky, které kniha neměla v době svého vzniku (např. razítko knihovny). Tento aspekt pozdějšího stavu knihy je něčím, od čeho běžní čtenáři dokáží na základě své znalostní základny odhlédnout, a tedy přidané prvky nepovažovat za součást originálu (knihy). Časovým ukotvením originálu pro čtenáře je tedy předpokládaný stav předtím (tj. bez razítka), zatímco pro knihovny je to stav v okamžiku zařazení knihy do fondu (tedy kniha již doplněná např. o razítko), od kterého se odvíjí jejich povinnost knihu uchovávat.

Než přikročíme k tematizaci originálu v digitálním světě, nejprve je třeba uvést podstatnou distinkci mezi dvěma stavy digitálního objektu: uložený / reprodukováný objekt. Reprodukováním zde rozumíme převedení uloženého digitálního objektu počítačovou technologií do lidsky vnímatelné podoby – zobrazení, přehrání apod. Toto užití pojmu je ve shodě s anglickým výrazem „rendering“, jak jej definoval Rothenberg (Rothenberg, 2000, s. 54) a později převzaly další texty, například standard PREMIS (2015, s. 272).

Z hlediska uloženého objektu nelze tvrdit, že originál v digitálním světě představuje zcela nový koncept, ačkoliv se tak často uvádí. Vytváření dokonalých (bitově identických) digitálních kopií má samozřejmě silnou analogii ve vydávání tištěných knih ve více výtiscích. Na některé fyzické objekty se navíc vztahuje i hledisko reprodukce (např. filmové pásy musejí být promítány). Hlavní rozdíl mezi fyzickými a digitálními objekty tak spočívá spíše ve vysoké míře komplexity digitálních technologií a jejich rychlému zastarávání, které ohrožuje možnosti adekvátní reprodukce. Následující část pojednává o originálu z hlediska uloženého objektu, otázkám reprodukce se věnuje poslední sekce.

Z hlediska časového ukotvení originálu můžeme zdroje rozdělit do tří skupin. První tematizuje originál ve smyslu fyzické předlohy digitalizace (Bearman & Trant, 1998; Caplan, 2008; Digital Library Federation, 2002; Digital Preservation Handbook, 2018; Deegan & Tanner, 2006; Federal Agencies Digitization Initiative, 2010; Fernie, 2008; Grycz, 2006; NISO, 2007). Klasická otázka věrnosti digitalizace se pojí právě s tímto ukotvením.

Druhá skupina textů směřuje originál do doby, kdy digitální objekt vznikl. V projektu InterPARES je „originál“ (*original*) definován jako „první úplný a účinný dokument“ (InterPARES project, 2005d, s. 5), resp. „originální dokument“ (*original record*) jako „první kompletní dokument, který má schopnost dosáhnout svého účelu (tj. je účinný)“ (InterPARES project, 2005a, s. 23). Časovým ukotvením je zde okamžik, odkdy může digitální dokument začít plnit svoji zamýšlenou funkci v rámci chodu organizace. Rothenberg zase poukazuje na potřebu zachovat původní funkcionalitu a vzhled (*original functionality, look, and feel*) originálního dokumentu (*original document*) (1999, s. 16). Tento požadavek se pojí s počítačovým prostředím, ve kterém byl digitální dokument vytvořen a které mělo podle původního záměru tvůrce sloužit k jeho reprodukování. Pro Rothenberga je klíčový zejména původní software (*original software*). Zachování původní funkcionality dokumentu podle něj umožní pouze emulace tohoto softwaru (Rothenberg, 1999, s. 1). Bearman (1999) v polemice s konceptem emulace jako jediného prostředku pro zachování původního dokumentu zdůrazňuje archivářský požadavek zdokumentovat původní kontext (*original context*) vytvoření dokumentu – ten podle něj sebelepší emulace naplnit nedokáže.

Třetí skupina zdrojů tematizuje originál ve fázi uchovávání v repozitáři. Dva zdroje [(Data Seal of Approval Board, 2016), (PREMIS, 2015)] označují originál za objekt, který do repozitáře dodal vkladatel. To však může, ale nemusí být dokument v podobě, v jaké byl původně vytvořen (vkladatel může původní objekt převést do jiného formátu, který je vhodnější pro archivaci). Rámec OAIS pak označuje balíček AIP první verze za originální balíček AIP (*the original AIP*) (ISO 2012a, s. 105). Z toho lze logicky vyvodit, že objekt CDO v něm obsažený je podle normy originálem. Opět však platí, že tímto originálním objektem CDO může, ale nemusí být to, co do repozitáře dodal vkladatel v balíčku SIP – repozitář může provést normalizaci objektu CDO do archivačního formátu). Skutečnost, zda dochází ke změně objektu CDO, by podle normy měla optimálně být zaznamenána v dohodě o dodávání dat (*submission agreement*), kterou by mezi sebou měl uzavřít repozitář a vkladatel (ISO 2012a, s. 36-37).

Pokus o vlastní definici digitálního originálu nalezneme jen u Rothenberga: digitální originál (*digital-original*) je podle něj „jakákoliv reprezentace digitální informační entity, u níž existuje nejvyšší možná pravděpodobnost, že zachovává všechny smysluplné a relevantní aspekty (*meaningful and relevant aspects*) této entity“ (Rothenberg, 2000, s. 66). Neurčitost této definice je záměrná, neboť podle Rothenberga je v současnosti obtížné všechny takové aspekty přesně specifikovat, a navíc je konkrétní specifikace závislá na dané oblasti užití dokumentu. Rothenberg tak bere v potaz fakt, že v digitálním světě je nemožné, aby každá digitální reprezentace udržovala všechny tyto aspekty (např. z důvodu zastarávání technologií). Tato definice není vztažena ke konkrétnímu časovému hledisku (digitální originál může být jakýkoliv objekt

v budoucnosti). Digitální objekt podle Rothenberga zahrnuje dva prvky: jak model (relevantní aspekty), tak všechny konkrétní objekty (z jakékoliv doby), které tomuto modelu odpovídají.

Domníváme se, že pokud máme užívat termín digitální originál, je nevhodné opouštět časové hledisko. Každý digitální objekt byl jednou vytvořen v určitém čase a v určitém (původním) počítačovém prostředí a jeho budoucí podoby (výstupy migrace, nového softwaru nebo emulace) již nikdy nebudou úplně stejné, jako byl objekt v době svého vzniku. Uvedený Rothenbergův koncept se vztahuje na všechny budoucí objekty jako model, ale zároveň nutně předpokládá, že tento model musel nějak vzniknout – a mohl vzniknout pouze z určitého původního (časově daného) objektu, který měl všechny (nejen relevantní) aspekty. Termín digitální originál tedy budeme užívat ve smyslu objektu CDO, který je v balíčku SIP určeném k dodání do repozitáře a který vznikl v určitý okamžik času pro specifické technologické prostředí.

Vzhledem k tomu, že digitální originál časem zanikne a zachovají se pouze některé jeho vlastnosti v nástupnickém objektu, a také k přihlídnutím k tomu, že digitální originál, který je digitalizátem knihy, má přímou vazbu na fyzický originál, navrhuje jako vhodnější výraz pro popis toho, k čemu se autenticita aktuálního objektu vždy nutně vztahuje, termín „referenční objekt“, který je samopopisný. Referenční objekt musí v případě digitalizátu knihy zahrnovat nejen digitální originál (časově první digitální objekt), ale také fyzickou předlohu (bez níž digitalizát nemůže existovat, protože část jejího informačního obsahu nutně reprezentuje) a rovněž specifikaci věrnosti, tedy bližší vyjádření vztahu mezi digitalizátem a předlohou.

### 3.3 Aspekt odvozeniny

Základní rámec pro zkoumání autenticity objektu jakožto uloženého objektu lze tedy popsat jako vztah mezi aktuálním objektem a referenčním objektem. Aktuální objekt je v tomto rámci vždy odvozeninou referenčního objektu, což platí i pro fyzické objekty (viz Gladneyho koncept pozdějšího stavu fyzického objektu jako „kopie“). Otázky autenticity aktuálního objektu tedy zahrnují posouzení autenticity odvozeniny na základě vztahu k referenčnímu objektu a udržení autenticity v průběhu procesu odvozování (řetězec odvozování). Aspekt odvozeniny podrobněji specifikují modely PREMIS a OAIS. PREMIS odlišuje dva typy digitálních odvozenin: odvozeniny vzniklé replikací (*replication*), které jsou bitově identické s předchozím objektem, a odvozeniny vzniklé transformací (*transformation*), které jsou bitově odlišné (PREMIS, 2015, s. 272-273). Model OAIS uvádí aspekt odvozování v souvislosti s celým balíčkem AIP. Všechna odvozování řadí pod kategorií „migrace“ (*migration*), což je archivační opatření prováděné repozitářem jako řešení problému zastarávání technologií (ISO 2012a, s. 22). Výraz „replikace“ ve standardu PREMIS tak odpovídá pojmům „renovace“ (*refreshing*) a „replikace“ (*replication*) v modelu OAIS, výraz „transformace“ (*transformation*) je v obou zdrojích užít ve shodném významu. Norma OAIS dále odlišuje vratnou (*reversible*) a nevratnou (*non-reversible*) transformací (ISO 2012a, s. 105-108). Příkladem jsou konverze nekomprimovaných obrazových dat do bezztrátové a ztrátové komprese. Nevratná

transformace logicky přináší vyšší riziko pro zachování autenticity, protože některé požadované vlastnosti mohou být nenávratně ztraceny. Také sama digitalizace je však formou transformace – běžnou digitalizací je vždy část informací nutně ztracena.

Požadavky na dlouhodobé uchování s ohledem na zachování autenticity znamenají zavedení vyššího stupně zabezpečení do běžných postupů a vyšších nároků na procesy odvozování, které jsou odborně stanoveny s ohledem na dlouhodobý horizont. V případě replikace to znamená užití mechanismů digitálního otisku (hašovací funkce), důkladné periodické kontroly, výběr vhodných datových nosičů a jejich včasnou výměnu. V případě formátových konverzí či jiných transformací je pro digitální archivaci jedním z jejích hlavních úskalí stanovit tyto procesy tak, aby nedošlo ke ztrátě požadovaných vlastností. Problémy transformace jsou shrnuty do následujících dvou oddílů (věrnost, klíčové vlastnosti).

### 3.4 Koncept věrnosti digitalizace

Koncept věrnosti je nepochybně nejrozšířenějším normativním konceptem v digitalizaci. Věrnost (*faithfulness*) předloze spojují přímo s autenticitou digitalizátu čtyři zdroje (Bearman & Trant, 1998; Federal Agencies Digitization Initiative, 2010, s. 51; Gladney, 2007, s. 97; Rothenberg, 2000, s. 52). Otázka věrnosti vyvstává v úplnosti až v okamžiku, kdy je vytvořen digitální originál. Proces digitalizace zahrnuje snímání předlohy, úpravy původních snímků (např. ořezy) a často převod do jiného formátu, než ve kterém byl uložen původní snímek (např. z TIFF do JP2). Konečná a detailní kontrola věrnosti výstupů digitalizace by tak měla probíhat až nad objektem CDO v balíčku SIP, a tato kontrola může znamenat, že ještě dojde opravám a bude vytvořen opravený balíček SIP. Proto má smysl považovat za digitální originál až bezchybný (zkontrolovaný) objekt CDO v balíčku SIP před jeho odevzdáním do repozitáře.

Směrnice DLF definuje věrné digitální reprodukce (*faithful digital reproduction*) jako „digitální objekty, které jsou optimálně formátovány a popsány s ohledem na jejich kvalitu (funkčnost a užitná hodnota), trvalost (dlouhodobé zpřístupňování) a interoperabilitu (např. napříč platformami a softwarovými prostředími)“ – jejich účelem je „přesně reprodukovat výchozí zdrojový dokument, s ohledem na jeho úplnost, vzhled původních stránek (včetně tonality a barvy) a správnou (tj. původní) posloupnost stránek.“ (Digital Library Federation, 2002, s. 2). Ke splnění tohoto cíle je uvedeno několik podmínek: obrazy vakátů (včetně zadní strany desek) budou začleněny jako části stanovené posloupnosti; bude možné propojit obecná deskriptivní metadata s digitálními součástmi objektu; bude možné z digitálních kopií reprodukovat (tiskem nebo online zobrazením) původní posloupnost částí, ale rovněž bude umožněna navigace mezi nimi (zobrazení následující strany, titulní strany apod.); digitální objekty bude možné citovat (Digital Library Federation, 2002, str. 4).

Grycz v souvislosti s digitalizací rukopisů uvádí: „Abychom zůstali věrni tradici výroby knih, měly by digitální obrazy vzácných tisků reprodukovat časem ověřené kvality původního (original) tištěného díla... Jedním z nejpatrnějších způsobů [jak toho docílit] je trvat na digitalizování knihy jako otevřené dvojstránky, bez ořezaných hran, tedy tak, aby čtenář mohl knihu vnímat, jako by spočívala na stole nebo

ji držel v rukou.“ (2006, s. 41). Tento způsob snímání je podle Grycze snímáním „v přirozeném stavu, který je věrný způsobu, jakým byla původní kniha navržena k prohlížení a užívání“ (Grycz, 2006, s. 42). Jedná se o specifický způsob digitální reprezentace tištěného dokumentu, nezvyklý pro masovou digitalizaci novodobých fondů.

To, s čím tematizace věrnosti implicitně pracují, je skutečnost, že knihy se v současné praxi digitalizují ve stavu, v jakém se nacházejí v době snímání, a účelem digitálního originálu v běžné praxi je věrně reprezentovat aktuální stav tištěné knihy. V případě, že by časovým určením fyzické předlohy byl zvolen stav knihy v době jejího vzniku, pak by fyzická předloha byla konceptuálním objektem, vytvořeným na základě výzkumu toho, jak mohla vypadat kniha v době svého vydání. Takováto digitalizace věrná původní podobě knihy by vyžadovala odborné digitální restaurování (např. odstranění zažloutlosti, vymazání razítka, zvýraznění vybledlého písma apod.). Tento přístup možná působí nezvykle v kontextu digitalizace knih, ale běžně se s ním setkáme u digitalizace filmů – viz například projekt Digitální restaurování českého filmového dědictví (<http://eea.nfa.cz/>) Národního filmového archivu. Spor o způsob restaurování, který tento archiv vedl s Asociací českých kameramanů, ukazuje na odlišné přístupy k určení autentické digitální reprezentace analogového filmu, přičemž jádrem sporu je právě specifikace referenčního objektu.

V případě digitalizace knih se však koncept věrnosti ještě komplikuje v tom, že knihu lze digitalizovat na různé úrovni abstrakce podle modelu FRBR. Ten pro popis dokumentu stanovuje čtyři úrovně abstrakce: dílo (*work*), vyjádření (*expression*), manifestace (*manifestation*) a jednotka (*unit*) (IFLA, 2009). Všechny úrovně (s výjimkou díla) lze poměrně dobře metodicky vztáhnout na knihu jakožto určitý typ dokumentu. V současné praxi masové digitalizace knih, kterou můžeme pojmenovat jako „rastrová digitalizace“ (předlohy jsou snímány s cílem reprezentovat vizuální vlastnosti knihy v rastrových obrazech) je zachycovaná úroveň abstrakce manifestace (vydání) a jednotka (exempláře) současně. Tato dvojí zachycovaná úroveň má analogii v katalogizaci tištěných knih. Z běžné praxe digitalizace však lze odvodit, že upřednostňovanou úrovní je vydání. Například pravidla pro poskytování dotací z podprogramu VISK7 vyžadují, aby nebyl digitalizován exemplář vydání, které již bylo digitalizováno: argument, že daná knihovna má jiný exemplář (téhož vydání), než který již byl zdigitalizován v nějaké předchozí digitalizaci, nelze uplatnit (Ministerstvo kultury, 2015, s. 6). Důraz na úroveň vydání platí i z hlediska zpřístupňování v digitálních knihovnách, kde primárně lze vyhledávat podle bibliografických údajů vztahujících se k vydání předlohy, neboť to je úroveň, která zajímá čtenáře především. Mohou se však vyskytovat projekty, které budou akcentovat specifičnosti exemplářů knih (např. poznámky vepsané do knihy jejich vlastníky, kteří jsou považováni za významné osobnosti). Pro zdůraznění úrovně exempláře je však potřeba specializovaný popis nad rámec běžné katalogizace knih z novodobých fondů.

Knihu však lze také digitalizovat knihu tak, že do digitální podoby se převede pouze její text se strukturálními vlastnostmi (odstavce apod.), bez znaků vydání (jakým je např. stránkování). Jinými slovy to znamená, že věrnost předloze lze pro digitalizaci specifikovat také na úrovni vyjádření. Příkladem je

digitalizace Shakespearových her Oxfordské univerzity (<http://ota.ox.ac.uk/desc/3014>). Objekt CDO je zde v textovém formátu TEI, který byl navržen právě pro zachycení strukturálních vlastností textu. Je také třeba si uvědomit, že takováto digitalizace neznamena ztrátovou transformaci (text v jeho struktuře je možné zdigitalizovat zcela bezztrátově) a že může být provedena zcela bez užití snímacích zařízení (tj. pracovníkem, který text knihy přepisuje a kóduje do formátu TEI). Tento způsob digitalizace je nezvyklý u novodobých knih, ačkoliv skrývá velký potenciál.

Z výše uvedené analýzy je zřejmé, že věrnost digitalizace není absolutní pojem. Je možné zvolit různé přístupy, které rovnocenným způsobem zajistí udržení věrnosti digitalizátu vůči předloze, ačkoliv každý povede k vytvoření odlišného digitalizátu. Věrnost však musí být předem (před zahájením digitalizace) specifikována.

Základní otázky věrnosti běžné rastrové digitalizace knih lze shrnout jako 1) jaký implementační model reprezentace zvolit (např. zda snímat knihu po jedné straně nebo dvojstránce, jakým způsobem provádět ořezy apod.) a 2) jaká konkrétní (technická) specifikace zajistí věrnou reprezentaci. Technická specifikace zahrnuje především dostatečné rozlišení, vhodnou bitovou hloubku a barevný model a adekvátní barevný profil a formát digitálního originálu. Formát musí být dostatečně kvalitní (např. bezztrátový) a z hlediska uchování navíc vhodný pro archivaci (např. s otevřenou dokumentací). Pokud je digitální originál již vytvořen v archivačním formátu, odpadá nutnost normalizace v repozitáři. Obzvláště důležitá je adekvátní barevná specifikace. Jedním z doporučených barevných profilů je Adobe RGB 1998 (Federal Agencies Digitization Initiative, 2010, s. 46). Zde je třeba dodat, že sám tento ICC profil má svoji oficiální (autentickou) podobu, kterou specifikovala firma Adobe a která nese název „Adobe RGB (1998) ICC profile“ (Adobe, 2005, s. 2). Jiné firmy si mohou vytvořit vlastní ICC profil pro namapování do barevného prostoru Adobe RGB 1998, ale nesmějí užívat tento název.

### 3.5 Klíčové vlastnosti

Klíčovými vlastnostmi zde označujeme kategorii vlastností, které mají být zachovány napříč řetězcem odvozování (zejména napříč transformacemi) objektu CDO. Norma ISO 14721 tuto vlastnost označuje jako „transformační vlastnost informací“ (*transformational information property*): „vlastnost informací, u které je uchování její hodnoty pokládáno za nutné, ale ne dostačující pro to, aby bylo možné ověřit, zda při jakékoliv nevratné transformaci byl zachován informační obsah“ (ISO 2012a, s. 27). Taková vlastnost informace je podle normy závislá na konkrétních interpretačních informacích (ISO 2012a, s. 106-107) a její zachování může „podstatným způsobem dokládat autenticitu“ (ISO 2012a, s. 24). Ve standardu PREMIS nalezneme analogický pojem „signifikantní vlastnosti“ (*significant properties*): „vlastnosti jednotlivého objektu, u nichž je subjektivně stanoveno, že je důležité je udržovat napříč archivačními opatřeními... i když mají objekty stejné technické vlastnosti, mohou se lišit vlastnostmi, které by měly být zachovány pro budoucí prezentaci nebo užití.“ (PREMIS, 2015, s. 50). Jak již bylo uvedeno, projekt InterPARES v podobném významu užívá spojení „podstatné aspekty“ (ovšem neuvádí žádnou definici)



a Rothenberg zase „smysluplné a relevantní aspekty“. Celkově lze shrnout, že koncept klíčových vlastností je přiznáním skutečnosti, že ve fázi uchovávání nelze v dlouhodobém horizontu zachovat všechny vlastnosti, které měl digitální originál, a to vzhledem k nevratnému zastarávání původních technologií (např. souborových formátů).

V rámci řízení celého životního cyklu digitalizátů je potřeba klíčové vlastnosti vztáhnout i na oblast digitalizace (tj. vlastnosti, které má digitalizace zachycovat) a jako cíl vytváření prezentačních variant. Prezentační varianty jsou specifické v tom, že v některých případech mohou být v nižší kvalitě (např. ztrátová komprese obrázků), aniž by ztrácely – z pohledu čtenářů – klíčové vlastnosti (rozdíl nemusí být vizuálně zaznamatelný); z hlediska potřeb archivace je však ztrátovost komprese problém a znamená ztrátu klíčových vlastností.

Klíčové vlastnosti lze celkově rozdělit na dvě vzájemně propojené skupiny: digitalizací převáděné vlastnosti fyzického originálu (např. vizuální vlastnosti a původní posloupnost stran nebo text ve své struktuře) a digitálně nativní vlastnosti (např. plnotextová prohlédatelnost), které jsou přidány v procesu digitalizace a které předloha nikdy neměla, protože jde o vlastnosti, které existují jen v digitálním světě.

Problémem je, jak už bylo uvedeno výše u Rotherbergova konceptu digitálního originálu, jakým způsobem určit, jaké všechny vlastnosti jsou důležité pro daný typ dokumentu, a tedy musejí být zachovány napříč transformacemi, a také to, jak tyto vlastnosti přesně specifikovat. Ve standardu PREMIS jsou pro zápis těchto vlastností dva opakovatelné elementy: typ signifikantní vlastnosti a její hodnota. Standard pouze uvádí, že signifikantní vlastnosti lze definovat na různých úrovních abstrakce (např. typ: „obsah“, hodnota „veškerý textový obsah a obrázky“; typ: „počet stran“, hodnota „7“) (PREMIS, 2015, s. 51) a žádné externí metadatové schéma pro popis jednotlivých signifikantních vlastností digitalizátů knih neexistuje. Přitom jde o zásadní otázku pro dlouhodobé uchovávání.

Jakkoliv je koncept signifikantních vlastností složitý a v současnosti nedostatečně specifikovaný, je nepochybně užitečný k tomu, aby ukázal na nutnost před každou transformací objektu CDO explicitně stanovit cíle, kterých má být dosaženo, a na základě těchto cílů výsledky transformace kontrolovat. Nezachování klíčových vlastností po provedení transformace pak lze jednoznačně považovat za narušení autenticity. Specifikace klíčových vlastností by měly být další součástí referenčního objektu, přičemž by měly být zdokumentovány před zahájením digitalizace jako její cíl digitalizace i jako cíl následného dlouhodobého uchovávání v repozitáři a cíl vytváření prezentačních variant v digitální knihovně. Například digitalizaci na úrovni vyjádření lze provést tak, že budou předlohy snímány černobíle a uloženy jako rastrové obrazy. Jelikož jejich klíčové vlastnosti budou pouze text v jeho struktuře, pak při budoucí migraci do formátu TEI bude postačovat přepis člověkem a ostatní informace mohou být ztraceny, protože nejsou podstatné. V tomto případě tedy černobílá digitalizace žádným způsobem nenarušuje autenticitu, protože byl zvolen jiný přístup.

### 3.6 Identita dokumentu

Koncept identity jako jedné ze složek autenticity detailněji rozpracovává projekt InterPARES, kde je identita celek charakterizujících a odlišujících atributů. Model OAIS uvádí obdobný koncept identifikačních informací (*reference information*), což jsou „informace, která jsou využívány jako identifikátor informačního obsahu“ (ISO, 2012a, s. 25). Uváděnými příklady z oblasti knihoven jsou perzistentní identifikátory a bibliografický popis. Deeganová a Tanner (2006) považují za jedno z hlavních rizik, které ohrožují autenticitu dokumentu, jeho chybnou identifikaci. Bearman a Trantová (Bearman & Trant, 1998) ve stejné souvislosti uvádějí problém existence různých verzí digitálních dokumentů nebo různých digitalizátů též fyzické předlohy, které jsou zpřístupňovány na internetu, ale nejsou dostatečně odlišeny; řešení spatřují v zavedení perzistentních identifikátorů. S perzistentními identifikátory spojuje identitu také Caplanová (2008). Gladneyho tvrzení, že nejsme ochotni považovat za autentický takový dokument, u něhož není známa identita jeho tvůrce, lze považovat za podkategorii identity dokumentu.

Tradiční definice autenticity jako vlastnosti dokumentu, který je tím, za co se vydává, souvisí právě s určením jeho identity. Tato definice vyvolává otázku, nakolik se nějaký objekt sám od sebe může za něco vydávat. To je nepochybně vždy záležitostí kontextu, ve kterém se objekt vyskytuje, sociokulturního prostředí a jím formovaných očekávání. V kontextu institucionalizované správy je však přesnější definice normy ISO 14721 (autenticita jako míra, do které osoba pokládá objekt za takový, za který *je vydáván* uchovávací organizací). Z toho vyplývá, že nezbytnou podmínkou je, aby organizace uchovávala spolu s dokumentem také jeho dokumentaci (metadata), přičemž dokumentace musí zahrnovat vyjádření identity dokumentu.

V případě tištěných knih znamená vyjádření identity primárně bibliografický záznam v katalogu. Problémy specifikace identity jsou dostatečně známy katalogizátorům, kteří zpracovávají knihy 19. století, jejichž produkce nebyla standardizovaná jako u současných knih. V kontextu obrovských knihovních fondů je navíc klíčovým prvkem vyjádření identity perzistentní identifikátor vydání knihy. Takovým je jistě identifikátor ISBN, ovšem tento systém neumožňuje přidělovat identifikátory zpětně (tj. před zavedením systému ISBN v dané zemi). Vhodným řešením může být zavedení systému kódů národní bibliografie (national bibliographic number, NBN), pro které je vyčleněno pole 015 ve standardu MARC21 a které mohou národní knihovny přidělovat záznamům své národní bibliografie (Hakala, 2001). Národní knihovny si však musejí vytvořit vlastní identifikační systém (tj. určit syntaxi a národní pravidla pro přidělování kódů NBN). Kódy NBN jsou již desítky let užívány například ve finské národní knihovně. V českém prostředí se však až donedávna neužívaly. Jako vhodné řešení způsobu perzistentní identifikace byly zavedeny až v roce 2010 (pod názvem „číslo ČNB“) na návrh autora tohoto článku (Cubr, 2010). Praktický problém s kódy NBN přirozeně spočívá v tom, že nejsou uvedeny v knize, pouze v záznamu, a že jsou přidělovány záznamům zpětně (to znamená, že musejí být při deduplikaci zachovány i kódy NBN zaniklých záznamů, jinak tento systém identifikace ztratí perzistenci). Perzistentní identifikátor

vydání knihy je v každém případě rozhodným prvkem pro určení identity knihy a pro potřeby správy knihovnických fondů i potřeby čtenářů.

Pro digitalizáty knih jsou z pohledu čtenáře rozhodující ty identifikační informace, které jsou uvedené v digitální knihovně spolu s prezentovaným dokumentem. Právě na základě nich je dokument knihovnou vydáván za to, co je zaznamenáno v těchto informacích. I zde sehrává rozhodnou roli perzistentní identifikátor digitálního dokumentu, od kterého se v internetovém světě navíc očekává, že bude pomocí něj možné dokument vždy vyhledat bez ohledu na změnu URL adresy. Takovou funkci nabízí několik mezinárodních identifikačních systémů (Handle, DOI, URN:NBN aj.), které vyžadují provoz dereferenční služby (resolveru). Tyto identifikační systémy také do určité míry samy obsahují určitou logiku specifikace identity dokumentu. Například ve finském systému URN:ISBN je tentýž identifikátor přidělen jakémukoliv digitalizátu knihy se stejným ISBN, tedy rozdíly mezi různými digitalizáty tento systém zaznamenat neumožňuje.

Celkově lze shrnout, že z hlediska potřeby udržení a zachování autenticity musí organizace při řízení životního cyklu (konkrétně před zahájením digitalizace) určit, jaké všechny atributy identity digitálního dokumentu musejí být vyjádřeny popisem tak, aby jednoznačně identifikovaly dokument, resp. referenční objekt. Návrh této specifikace obsahuje čtvrtá kapitola.

Z autorovy praxe se současnou digitalizací knih v českém prostředí je nutno poznamenat, že jeden z hlavních problémů identity dokumentu, a tedy zdroj rizik pro jeho autenticitu, jsou opravy digitalizátů knih, které se vykonávají nikoliv při kontrole balíčku SIP, ale až po jeho vystavení čtenářům v digitální knihovně. Nejzávažnější pozorované chyby lze typologizovat následovně: neúplné snímky (oprava: doskenování); nekvalitní snímky (oprava: opětovné skenování); strukturální změna dokumentu: rozdělení dokumentu na více dokumentů nebo spojení více dokumentů v jeden dokument; nesoulad mezi dokumentem a přidruženými bibliografickými metadaty (bibliografický popis je popisem jiného digitalizátu, a tak je dokument v digitální knihovně vydáván za něco, čím není);

Koncept identity jako celku charakterizujících a odlišujících atributů je však potřeba rozšířit i na úroveň objektu CDO – jako digitální identitu. Digitální identitu lze rozdělit do dvou skupin: konkrétní digitální identita a typová digitální identita. Typová digitální identita znamená specifikaci určitého prvku, parametru, vlastnosti nebo funkce digitálního objektu, který lze aplikovat na více objektů (např. formát, rozlišení apod.). Konkrétní digitální identita je souhrn specifických atributů konkrétního (na konkrétním nosiči uloženého) digitálního objektu (např. velikost souboru, pixelová velikost). Digitální identita je něco, co se bude v průběhu času nutně měnit (např. formátovou migrací vznikne nový objekt CDO, který bude mít jinou digitální identitu), ale v mezidobí musí být udržována stejná.

Digitální identitu je potřeba před každou transformací (vč. digitalizace) také předem specifikovat. To na první pohled může působit triviálně (jakýkoliv digitální objekt musí být předem specifikován, jinak by nemohl vzniknout), nicméně v praxi je řada specifikací dána implicitně (např. užitím konkrétního softwaru nebo jeho výchozím nastavením) nebo není stanovena nedostatečně podrobně. Typickým příkladem

v tomto ohledu je formát. Formát nelze identifikovat jen jeho názvem (např. „PDF“) – důležitá je také jeho verze nebo podtyp (např. „PDF/A 2b“). Jediný existující systém pro podrobnou identifikaci je PUID, což je identifikátor interpretačních informací (nejen formátu) (Brown, 2006, s. 4). Například PUID pro JPEG verze 1.00 je „fmt/42“, pro verzi 1.01 „fmt/43“ a pro verzi 1.02 „fmt/44“. Široce užívaný registr MIME (<http://www.iana.org/assignments/media-types/media-types.xhtml>) odlišuje formáty jen na základě typu a názvu. Tak například formát JPEG všech verzí má podle MIME označení „image/jpeg“. Systém PUID však nepokrývá všechny existující formáty a způsob doplňování jeho záznamů není zcela precizní. Dalším příkladem atributu digitální identity, který musí být pečlivě vybrán a precizně specifikován, je barevný profil. Bez jeho vhodné volby může dojít k tomu, že obrazová data budou snímána nebo uložena v nevhodném proprietárním ICC profilu a v důsledku toho hrozí, že budou špatně zobrazena.

### 3.7 Integrita dokumentu

Pojem integrity je ve zkoumaných zdrojích užíván poměrně různorodě. Ve standardu PREMIS je základem pro posouzení autenticity objektu možnost ověřit integritu jeho zdroje (PREMIS, 2015, s. 267). Ve směrnici DSA jsou autenticita a integrita pojímány jako dvě odlišné kategorie stejné úrovně (Data Seal of Approval Board, 2016). V prvním uvedeném archivářském projektu je integrita celkem spolehlivosti a autenticity; v projektu InterPARES tvoří integrita s identitou dvě základní složky autenticity a je definována jako úplnost a neporušenost ve všech podstatných aspektech, což propojuje pojem integrity s uvedeným konceptem klíčových vlastností. V modelu OAIS pojem integrita není definován. Související pojem neporušenost (*fixity*) znamená zachování bitové totožnosti (ISO, 2012a, s. 22). V modelu OAIS by tedy širšímu pojetí integrity projektu InterPARES zhruba odpovídalo spojení pojmů neporušenost (pro replikace) a transformační vlastnost informací (pro transformace).

Z koncepcí identity a integrity vyplývá, že posuzování integrity je založeno na znalosti identity objektu – nejprve musí být známo a přesně popsáno, o jaký objekt jde (tedy musí být vyjádřena jeho identita), a teprve poté lze posuzovat, zda nebyla porušena jeho integrita. Otázka autenticity objektu, o kterém nic nevíme, je tedy sama o sobě problematická. Pro oblast institucionalizovaného uchovávání lze tedy říci, že sama knihovna by měla vyvinout maximální úsilí při zjišťování (ve fázi akvizice), stanovování (ve fázi vlastní produkce) a popisu (po vytvoření dokumentu) identity digitálního dokumentu, který musí zahrnovat i popis klíčových vlastností. Určení digitální identity pak znamená přesně specifikovat parametry objektu CDO a mít k dispozici validační nástroje, které ověří splnění požadovaných parametrů po dokončení procesu odvozování. A zde se skrývá jeden z velkých problémů digitální archivace. I sám formát objektu CDO totiž musí být autentický – takový, který je plně v souladu se svou oficiální specifikací. Soubor v chybně vytvořeném formátu se technicky vydává za něco jiného, než čím je. Mezi další problémy digitální identity patří následující: ne každý softwarový nástroj dokáže vytvořit autentický formát; různé verze formátového validátoru mohou přinášet rozdílné výsledky; různé verze nástroje pro formátovou identifikaci pomocí identifikátoru PUID mohou odlišně určit formát; objekt ve formátu, který

nemá otevřenou a dostupnou specifikaci, de facto nelze transparentně validovat, a tedy nelze ověřit jeho autenticitu.

### 3.8 Provenience

Koncept provenience se ve zkoumaných zdrojích definičně liší, ačkoliv spíše v detailech. Provenience je popisována jako: dílčí typ konstituentu kontextu, který je předmětem archivního popisu konkrétní archivní sbírky (někdy tříděné podle původce) (InterPARES project, 2005b); historie událostí vztahujících se k digitálnímu objektu (Caplan, 2008, s. 8); původ objektu a historie jeho změn (NISO, 2007, s. 55); historie objektu (PREMIS, 2015, s. 2); informace o tvůrci a události vzniku a historii odvozování (Gladney, 2007, s. 97). Již uvedená definice normy ISO 14721 de facto v nejobecnější rovině zahrnuje vše výše uvedené (původ / zdroj informačního obsahu, všechny jeho změny a informace o všech správcích). Podle normy je odpovědností archivu OAIS vytvářet a uchovávat provenienční informace od okamžiku příjmu balíčku SIP od vkladatele; informace z předchozí doby by však měl dodat vkladatel; součástí provenienčních informací je podle normy také popis transformační vlastnosti informací (ISO, 2012a).

Je zřejmé, že provenienční informace musejí být zaznamenány v metadatech a dokumentaci a musejí být průběžně doplňovány. V tom se liší otázky posuzování autenticity institucionalizovaně spravovaných dokumentů a objektů mimo tento kontext, kdy se nelze opírat o to, že k objektu existuje dokumentace jeho historie.

### 3.9 Dokumentace

Nutnost získávání a udržování dodatečných informací (např. ve formě metadat), které slouží jako prostředek pro udržení a posuzování autenticity, uvádí řada tematizací (Bearman & Trant, 1998; Caplan, 2008; Data Seal of Approval Board, 2016; Digital Preservation Handbook; 2018; Gladney, 2007; NISO, 2007; PREMIS, 2015; UNESCO, 2016). Dvě z nich navíc zdůrazňují, že sama metadata musejí být autentická, k čemuž slouží „metametadata“, mezi něž patří například informace o zdroji metadat a způsobu jejich vytváření (NISO, 2007, s. 85), (UNESCO, 2016, s. 15).

Z hlediska autenticity jsou metadata a další dodatečné informace, které spolu s dokumentem musí uchovávat paměťová instituce, de facto jediným způsobem, jak zaznamenat provenienční informace a klíčové vlastnosti. Metadata jsou rovněž důležitá pro záznam informací o identitě (část těchto informací je také obsažena v samotném digitalizátu, např. ve snímku tiráže knihy apod.). V požadavcích na udržování autenticity v projektu InterPARES je uvedena nutnost vytváření a udržování dokumentace procesů odvozování a jejich důsledků. Tento požadavek zahrnuje i nutnost zpřístupňovat uživatelům informace o procesu, jehož důsledkem odvozenina nepřenáší všechny vlastnosti původního objektu v úplnosti (InterPARES project, 2005c, s. 11). Norma ISO 16363 pro certifikaci důvěryhodnosti repozitářů uvádí mj. nutnost dokumentovat pravidla uchovávání (popis toho, jakým způsobem jsou dokumenty uchovávány), dílčí postupy v repozitáři a záznamy o provedených archivačních opatřeních

(ISO, 2012b). Do posledně jmenované kategorie spadají jak protokoly vytvářené pracovníky repozitáře, tak i záznamy aplikací.

Základním rámcem pro metadata a dokumentaci jsou koncepty interpretačních a archivačních informací modelu OAIS. Je potřeba upozornit, že metadata zaznamenávající tyto typy informací jsou podle tohoto modelu stále „jen“ digitální data, tj. je potřeba, aby k nim také existovaly interpretační informace, což v principu vede k nekonečné rekurzivitě. Tato rekurzivita může být ukončena pouze tehdy, pokud jsou na konci řetězce dalších interpretačních informací takové interpretační informace, které jsou zaznamenány v podobě fyzického objektu (např. tištěná dokumentace) – takový objekt je již vnímatelný člověkem bez potřeby počítačového zprostředkování (ISO, 2012a). Tento problém však není v současnosti závažný, neboť digitální metadata jsou v běžné praxi zaznamenávána v textových formátech (typicky v XML), jejichž interpretace počítačovým prostředím se nepovažuje za problematickou, a to ani z dlouhodobého hlediska.

Metadatové standardy jsou nutně determinující prvek. Standard PREMIS např. „vnucuje“ určitou logiku zápisu metadat. Změna objektu znamená vždy vytvoření nového objektu a nutnost zaznamenat ji jako vztah mezi zdrojovým a cílovým objektem (PREMIS, 2015, s. 265). Jinými slovy, standard nepřipouští koncept verzování objektu. Logika vepsaná v mezinárodních metadatových standardech je racionální a užitečná jako základní instrukce pro vytváření metadatových záznamů pro knihovny, které nemají kapacity na to porozumět všem požadavkům životního cyklu z hlediska autenticity. Možnosti popisu digitální identity pak vymezují zejména metadatové standardy pro technická metadata pro určitý typ objektu (např. MIX pro obrazová data).

Řídit se mezinárodními (nebo široce užívanými) standardy znamená záruku mnohem vyšší interoperability. Zároveň to vyžaduje vyčkávat na případné doplnění metadatových prvků pro něco, co zatím popsat nelze, v budoucí nové verzi standardu či zcela novém standardu. Příkladem významné změny metadatového standardu je vydání třetí verze PREMIS (2015), které umožňuje zápis událostí (jako součást provenience) na intelektuální entitě (například pro záznam změny bibliografických metadat). Vytvářet vlastní metadatový standard je intelektuálně náročné a přináší riziko nižší interoperability i nejistotu vývoje. Vytvářet externí schémata (jako rozšíření stávajících standardů) je smysluplnější.

Existující užívané standardy jsou dostatečným základem pro současnou praxi digitalizaci tištěných knih, ovšem s tím, že vyžadují správný způsob převzetí do aplikačních metadatových profilů. Některé informace je však obtížné zaznamenat do metadat a je vhodnější je zaznamenat do dokumentace – příkladem jsou klíčové vlastnosti, které je nejlepší popsat v přidružené dokumentaci digitalizačního projektu a v budoucnosti je případně převést do podrobnějších elementů (v době, kdy budou dostupné například ve standardu PREMIS).

Do oblasti autenticity metadat spadá mj. problém toho, jakým způsobem jsou metadata získávána. V případě technických metadat je důležité, aby v maximální možné míře byla získávána důvěryhodnými

nástroji pro extrakci metadat přímo ze souborů, nikoliv z přednastavených hodnot, kdy vždy hrozí lidská chyba (např. při výměně skeneru se neprovede změna názvu skeneru). V případě bibliografických metadat je základním požadavkem, aby digitalizaci předcházela pečlivá kontrola záznamu nebo rekatalogizace – požadavek, který tak často v české praxi zůstává nenaplněn.

### 3.10 Reprodukce počítačovou technologií

Úvodem tohoto oddílu je třeba zmínit skutečnost, že ne všechny digitální objekty jsou určeny k reprodukování lidskému uživateli. Distinkci na reprodukované objekty (*rendered objects*) a nereprodukované objekty (*non-rendered objects*) sice nenalezneme v normě ISO 14721, ale je uvedena ve výkladové publikaci (Giaretta, 2011, s. 33-37). Nereprodukované digitální objekty jsou určeny ke zpracování jinými počítačovými systémy (např. datové sady vědeckých archivů obsahující obrovské množství údajů zaznamenaných z pozorování). Digitalizáty knihy jako typ reprodukováných objektů jsou určeny k tomu, aby byly softwarovými nástroji zobrazovány čtenářům.

Koncept interpretačních informací modelu OAIS zahrnuje aspekt reprodukování. Interpretační informace jsou potřebné k tomu, aby byl objekt CDO převeden do smysluplného celku, což v případě některých digitálních objektů nutně znamená reprodukování. To je nejvíce patrné na popisu dvou typů softwarových nástrojů pro interpretační informace (ISO 2012a, s. 68-69). Software pro zobrazení interpretačních informací (*representation rendering software*) je software, který umožňuje interpretační informace reprodukovat v podobě srozumitelné lidem. Příkladem je prohlížeč Adobe Acrobat, který dokáže zobrazit formátovou specifikaci uloženou v PDF. Zpřístupňovací software (*access software*) je software, který dokáže prezentovat samotný informační obsah (nebo jeho část). Tímto případem je nepochybně také sama digitální knihovna jako celek – zde se prezentují digitalizáty uživatelům. Jinými slovy, software pro zobrazení interpretačních informací reprodukuje interpretační informace a slouží jako pomůcka pro uchovávání informačního obsahu, zatímco zpřístupňovací software reprodukuje samotný informační obsah a slouží jako prostředek pro zpřístupňování informačního obsahu cílové komunitě. Oba typy softwaru jsou současně samy typem interpretačních informací a jsou součástí sítě interpretačních informací (*representation network*). Norma však v této souvislosti varuje, aby se repozitář nespoléhal pouze na zpřístupňovací software – v případě zastarání tohoto softwaru se informační obsah stane nereprodukovatelným, a bude tedy nenávratně ztracen.

Projekt InterPARES k otázce reprodukce uvádí: „Mezi digitální reprezentací dokumentu v úložišti a formou, ve které je prezentován k užití, nutně existuje zásadní rozdíl. Vždy je nutné využít nějaký software k převedení uložených digitálních bitů do dokumentové formy archiválie. To nevyhnutelně vytváří rizika, že bez ohledu na to, jak dobře byla digitální data zabezpečena na úložišti, může být archiválie nevhodně pozměněna v procesu načítání uložených bitů a prezentování uživatelům... proces uchovávání elektronické archiválie přesahuje rámec zabezpečeného uložení... začíná prvotním aktem uložení dokumentu a pokračuje jeho reprodukováním.“ (InterPARES project, 2005b, s. 5).

Oba stěžejní modely upozorňují na základní problém digitální archivace – potřebu trvale udržitelné adekvátní reprodukce. Nicméně autenticitu spojují pouze s uloženým objektem. Podobně i Caplanová uvádí, že zajištění reprodukovatelnosti (*renderability*) je „srdcem procesu digitální archivace“, nicméně objekt podle ní může být autentický, i když jej nelze reprodukovat (2008, s. 9).

Autenticitu ve spojení s reprodukováním nalezneme explicitně pouze u Grycze – podle něj by digitalizát měl „zachytit dostatečné informace, které umožní badateli (třeba za 300 let od současnosti) vidět a reprodukovat digitalizované obrazy způsobem, který by konstituoval autentickou reprezentaci toho, jak originál vypadal v době jeho snímání.“ (Grycz, 2006, s. 47).

Je to však právě aspekt autentické reprodukce, který je výchozím bodem pro posuzování autenticity. Autenticita je tedy primárně vlastnost zobrazeného digitalizátu v digitální knihovně, nikoliv uloženého objektu, jakkoliv jsou bezpečné uložení a adekvátní specifikace objektu CDO fyzicky i logicky nezbytnou podmínkou pro jeho reprodukci.

Každý akt reprodukce objektu CDO, jak ostatně vyplývá ze samotného slova, je znovuvytvořením objektu v daný okamžik v daném softwarovém systému a ve specifických podmínkách koncového zařízení čtenáře (kvalita monitoru, okolní světlo apod.). Pokud odhlédneme od konkrétních podmínek vnímatele, je rozhodným aspektem, který je relativně měřitelný, adekvátní zpřístupňovací software. Zde hlavní problém představuje zastarávání softwaru. Z tohoto důvodu musí být software digitální knihovny trvale vyvíjen v závislosti na vývoji internetových prohlížečů, požadavcích čtenářů a změnách formátu. Pro údržbu digitální knihovny je nutné vybírat a testovat nástroje, které umožní reprodukovat informační obsah bezchybně a v souladu s klíčovými vlastnostmi, a v delším časovém horizontu je také potřeba vybírat vhodné prezentační formáty (v současnosti hojně užívaný JPEG jednou také zastará). Dalším problémem je však (mnohem méně reflektovaná) skutečnost, že i různé současné softwarové aplikace mohou tentýž objekt CDO interpretovat různě. Žádná z uvedených tematizací se nezabývá tímto problémem. V praxi přitom nejde o výjimečný jev, byl popsán například u formátu PDF (British Library, 2015, s. 3-5) nebo rastrových formátů [viz např. (Vychodil, 2013, s. 98-108)]. Do určité míry jde jistě o otázku kvality softwaru. Částečně však může jít také o problém kvality dokumentace (tj. zda je daný formát dostatečně podrobně a jednoznačně popsán tak, aby bylo na základě této dokumentace možné naprogramovat adekvátní zpřístupňující software). V případě komplexních formátů pak může být otázkou sama realizovatelnost tak jednoznačného popisu, který by vylučoval odlišné interpretace při vytváření softwaru.

Reprodukce v digitální knihovně je neodlučně spojena s otázkou internetového zpřístupňování. Směrnice MINERVA v souvislosti s prezentováním výstupu digitalizačního projektu veřejnosti uvádí: „Název domény tvoří část „obchodní značky“ projektu a napomůže koncovým uživatelům určit autenticitu poskytovaného obsahu. Názvy domén by tedy měly být jasně označeny buď názvem projektu, nebo organizací, která projekt zajišťuje. Pro zvýšení důvěry koncových uživatelů, že si vyměňují informace se správným webovým sídlem projektu, může být někdy vhodné zabezpečit síťové spojení mezi klientem a serverem protokolem Secure Sockets Layer (SSL).“ (Ferne, 2008, s. 66). Tato směrnice tedy autenticitu



dokumentu pojímá z hlediska identity jeho původce: důvěryhodnost webového sídla digitální knihovny (tj. skutečnost, že tato digitální knihovna je skutečně provozována původcem, za jehož webové sídlo se vydává) plošně činí zde prezentované digitální dokumenty autentickými. Standard PREMIS v souvislosti s původcem zpřístupňovaného obsahu uvádí užití digitálního podpisu. Repozitář může připojit digitální podpis k datům, která vydává jako balíčky DIP, jako doklad tohoto, že tento archiv je skutečně původcem daných dat (PREMIS, 2015, s. 259). Lze říci, že v případě, kdy digitální knihovna umožňuje čtenáři stáhnout si celý digitalizát knihy, může být vhodné současně s ním zpřístupňovat také jeho digitální otisk, a to pro ověření, že při stahování nedošlo k bitové změně.

Dalším aspektem internetového zpřístupňování je citování – požadavek, že „cokoliv je citováno, je stejné, jako bylo v době svého vzniku, pokud nejsou v doprovodných metadatech uváděny žádné změny.“ (Digital Preservation Handbook, 2018). Tento poslední aspekt opět ukazuje na potřebu spolu se zpřístupňovaným (reprodukováným) dokumentem prezentovat také dostatečnou dokumentaci. Tento požadavek tvoří základ požadavků na autenticitu z hlediska čtenáře. Digitalizát knihy musí být zobrazen spolu s metadaty a odkazy na dokumentaci, které popisují všechny nezbytné atributy jeho identity. Tyto atributy musejí být vyjádřeny v úplnosti a bezchybně, aby čtenář nebyl uveden v omyl. Tento aspekt lze ze zkušenosti z české digitalizační praxe považovat za jeden z hlavních problémů autenticity: pokud je prezentován špatně popsáný dokument, je autenticita narušena z hlediska toho, za co je dokument vydáván. V současné české praxi jsou, jak bylo uvedeno, velkým problémem pro autenticitu pozdější opravy dokumentu, který ale již byl prezentován čtenářům v digitální knihovně a mohl být citován, což znamená, že čtenář pracoval s něčím jiným, než za co dokument považoval. V digitalizátu knihy mohly například chybět stránky, které nebyly naskenovány, ale čtenář si tohoto faktu nevšiml. Tato rizika narušují důvěryhodnost nejen digitální knihovny, ale také výzkumné práce s prameny.

#### 4 Základní rámec pro konceptualizaci autenticity digitalizátů knih

Přehled konceptuálních prvků spojených s autenticitou v předchozí kapitole nám umožňuje vytvořit následující redefinici autenticity:

***autenticita je vlastnost (aktuálního) digitálního dokumentu, který je takovou odvozeninou od stanoveného referenčního objektu, která zachovává jeho klíčové vlastnosti při uložení i reprodukci počítačovou technologií, a jehož identita a provenience jsou dostatečně zdokumentovány (včetně popisu referenčního objektu a řetězce odvozování a správy).***

Tato definice není v rozporu se zkoumanými zdroji (s výjimkou Caplanové, podle níž může být objekt autentickým, ačkoliv jej není možné reprodukovat), nýbrž shrnuje a zpřesňuje různé aspekty v nich uvedené do podoby komplexního konceptu autenticity. Primárně rozvíjí Gladneyho definici (která nezahrnuje aspekt reprodukce). Navržená definice také předpokládá, že referenční objekt autenticity je předem specifikován. Tato definice může být vztažena na jakýkoliv typ digitálního dokumentu, který je

určen k reprodukování počítačovou technologií, a to za předpokladu, že organizace může řídit celý jeho životní cyklus.

V úvodu představená koncepce MacNeilové a Makové pojímá autenticitu jako sociokulturní konstrukci a připouští legitimitu odlišných perspektiv k jejímu stanovování. V kontextu řízení celého životního cyklu dokumentu jednou organizací ovšem platí, že je to právě tato organizace, která může (a měla by) konstruovat model autenticity, který bude aplikován při digitalizaci a následném dlouhodobém uchovávání a zpřístupňování čtenářům. Základem tohoto modelu je právě stanovení referenčního objektu. Je samozřejmé, že objektivně vždy musí existovat nějaká fyzická předloha a nějaký její digitalizát, aby bylo možno vůbec hovořit o autenticitě digitalizátu knihy (a konstrukce tohoto modelu je tedy do určité míry vždy predeterminována), nicméně specifikace vztahu mezi nimi, a tím i stanovení rozmezí, ve kterém se mohou následné otázky udržení a posouzení autenticity pohybovat, do značné míry záleží také na přístupu, který knihovna zvolí při digitalizaci.

Účelem navrhované konceptualizace je poskytnout rámec, ve kterém by se měla tato specifikace referenčního objektu pohybovat tak, aby bylo možno měřitelným a důvěryhodným způsobem deklarovat udržení autenticity a umožnit její posouzení, a to jak odbornými pracovníky repozitáře, tak čtenáři (členy cílové komunity). Precizní specifikace referenčního objektu je důležitá jednak z dlouhodobého horizontu, kdy informace o kontextu digitalizace již nebudou dostupné a pravděpodobně již nebudou existovat fyzické předlohy. Z tohoto důvodu sehrává zásadní roli systematicky sepsaná dokumentace digitalizačního projektu. Tato specifikace je také důležitá jako projev transparentnosti produkce digitalizátů (a následného dlouhodobého uchovávání a zpřístupňování) vůči cílové komunitě, tedy čtenářské obci. Znalost kontextu digitalizace (uvedeného v dokumentaci) také umožní porozumět původním technologickým volbám, které byly provedeny v době digitalizace. Starší digitalizace například čelily větším technologickým omezením (např. nízké rozlišení skeneru) než současná produkce, zároveň také při každé digitalizaci hraje roli finanční, časové a organizační hledisko, které do značné míry omezuje digitalizační výstupy.

Jak bylo argumentováno v předchozí kapitole, digitalizaci lze provádět pro různé účely a velmi odlišnými způsoby. Ostatně to ukazují i příklady projektů, jako jsou Gutenberg, digitalizace firmy Google, masové rastrové digitalizace národních knihoven nebo menší digitalizace zaměřené na vyšší kvalitu výstupů (např. digitalizace do formátu TEI). Podrobná dokumentace digitalizačního projektu tedy pomůže osvětlit účel digitalizace a zohlednit cílovou komunitu, pro niž je určena. Na základě dokumentace pak má čtenář možnost se dozvědět, co od dané digitalizace může / nemůže očekávat. Výchozím a rozhodujícím prvkem je v tomto ohledu, jak již bylo uvedeno, specifikace referenčního objektu autenticity.

#### **4.1 Specifikace referenčního objektu autenticity**

Referenční objekt v případě digitalizátu knihy musí zahrnovat tyto složky:

- a) fyzickou předlohu, a to ve vybraném časovém ukotvení (současný stav / předpokládaný původní stav knihy v době jejího vzniku);

- b) digitální originál, který je objektem CDO v balíčku SIP;
- c) klíčové vlastnosti, které musejí být udržovány i v budoucnosti;
- d) specifikaci věrnosti, tedy explicitní vyjádření vztahu mezi fyzickou předlohou a digitálním originálem.

Digitální originál je logicky vždy určitou reprezentací fyzické předlohy, ale tuto reprezentaci musí vymezovat explicitní specifikace věrnosti. Autenticita se tedy nutně vztahuje jak k fyzické předloze (která musí být vždy nějak zachycena digitalizací), tak k digitálnímu originálu, který je původním digitálním objektem a jehož klíčové vlastnosti musejí být udržovány v průběhu jeho uchovávání a zpřístupňování čtenářům v dlouhodobém horizontu.

Specifikace věrnosti musí být prvním krokem pro definování referenčního objektu autenticity. Určení věrnosti zahrnuje dva prvky – stanovení úrovně abstrakce podle FRBR a volba metody odvozování při produkci (která začíná snímáním tištěné předlohy a končí vytvořením objektu CDO v balíčku SIP).

Nejprve je nutno určit, na jaké úrovni abstrakce podle FRBR má být předloha snímána, což automaticky determinuje stanovení klíčových vlastností a má vliv na časové určení fyzického originálu a následnou volbu objektu CDO i výběr metody odvozování.

Na úrovni vyjádření jde v případě knih o textové vyjádření (odhlédneme-li od specifických případů, např. když jsou součástí vyjádření vlastní spisovatelovy ilustrace). Klíčovou vlastností musí být textový obsah ve své struktuře. Otázka, jak vyjádřit všechny strukturální vlastnosti textu knihy, může být komplexní analytický úkol. Tento úkol však do značné míry a již dlouhou dobu řeší konsorcium vyvíjející standard TEI. Doporučeným postupem je tedy vyjít z tohoto standardu, neboť ten v současnosti poskytuje velké množství elementů, kterými lze popsat různé strukturální vlastnosti textu obecně i specifické vlastnosti různých žánrů (próza, poezie, divadelní hra aj.).

V případě digitalizace na úrovni vyjádření může být optimálním řešením, aby digitální originál tvořila jediná komponenta – textová ve formátu TEI XML. Co se týče časového určení fyzické předlohy, na této úrovni abstrakce musí jít nutně o původní stav (účelem je převést text knihy v úplnosti a bezchybně), jehož podoba však v případě čitelných úplných textů nemusí být předmětem výzkumu.

Následně musí být stanovena a předem v dokumentaci zaznamenána zvolená metoda odvozování. Digitalizaci lze provést tak, že specialista přepíše text knihy přímo do formátu TEI, tedy se zakódováním strukturálních vlastností. Nebo je možné postupovat tak, že kniha bude snímána jako obrazová data, na kterých pak bude provedeno OCR a výsledný text specialista jen zreviduje a zakóduje do formátu TEI. Způsoby odvozování tedy mohou být různé, ale musejí být pečlivě zvoleny a popsány. Při posuzování autenticity při tomto typu digitalizace může hrát roli i to, jak posuzovatel zhodnotí, která z uvedených metod odvozování je spolehlivější z hlediska věrného převodu textového obsahu knihy. Při tomto posuzování může být důležitá informace o praktické realizaci projektu (např. poměr počtu pracovníků k počtu digitalizovaných knih nebo úroveň jejich zaškolení v kódování textu do formátu TEI). Konkrétní

pracovní postup by proto měl být také řádně zdokumentován. V každém případě platí, že jakákoliv chyba v textu znamená narušení věrnosti předloze, a tedy i autenticity tohoto typu digitalizátu knihy. Text tedy musí být převeden v úplnosti a v celé své struktuře, aby bylo možné konstatovat plné zachování autenticity při převodu do digitálního originálu.

Cílem digitalizace na úrovni manifestace (vydání) / jednotky (exempláře) je věrná vizuální reprezentace knihy. V případě úrovně vydání může jít buď o reprezentaci jejího současného stavu, nebo (předpokládaného) původního stavu. V případě úrovně exempláře logicky pouze o reprezentaci současného stavu (jde o zachycení prvků přidaných do knihy po jejím vzniku, např. známou osobností). Způsob zachycení současného stavu je v odborné literatuře poměrně dobře popsán. V druhém případě musí být užita metoda digitálního restaurování; jakkoliv takový přístup působí nezvykle, existuje možnost jej aplikovat. V každém případě i informace o tom, že je zachycován současný stav knihy, by měla být explicitně vyjádřena v dokumentaci digitalizačního projektu, protože je možné, že v budoucnosti nebude tento běžný úzus již znám a budoucím badatelům takové informace mohou pomoci k porozumění kontextu digitalizace. Celkově lze říci, že rozdíl mezi digitalizací na úrovni vydání a exempláře knihy spočívá především v míře specifikace identity dokumentu (to znamená, že je digitalizace na úrovni exempláře vyžaduje podrobnější popis).

Klíčové vlastnosti pro digitalizaci na úrovni vydání / exempláře lze v obou případech stanovit alespoň ve velmi obecné rovině. Navrhované základní hodnoty jsou: plná barevnost, dobrá čitelnost textu, původní posloupnost stran, úplnost stran a listovatelnost (tento výčet není úplný). Tyto vlastnosti musejí být převedeny do odpovídajících parametrů digitální identity objektu CDO (např. plná barevnost vyžaduje bitovou hloubku 24 bitů a vhodný barevný model). Klíčové vlastnosti však mohou obsahovat i speciální digitálně nativní vlastnosti, jako je plnotextová prohledatelnost a kopírovatelnost textu. Aby digitální originál splňoval všechny tyto vlastnosti, je objekt CDO v současné praxi tvořen jak obrazovou, tak textovou komponentou (typicky výstupy z OCR), přičemž musí být možnost obě komponenty v určité softwarové aplikaci vzájemně propojit, aby bylo možné „prohledávat obraz“. Je zřejmé, že užitím současné metody odvozování textové komponenty (proces OCR bez lidské kontroly) nelze zaručit stoprocentní přesnost, a tedy vlastnost plnotextové prohledatelnosti nemůže být naplněna beze zbytku. Optimálně by tedy měla být tato klíčová vlastnost zpřesněna, např. „plnotextová prohledatelnost – minimálně z 95 %“. Pokud by však referenční objekt ve své specifikaci nezahrnoval tyto digitálně nativní vlastnosti, pak by objekt CDO mohl být tvořen pouze obrazovou a strukturální komponentou – míra autenticity by tím nebyla nijak snížena, protože by tuto podobu popisovala (a tím určovala) právě specifikace referenčního objektu.

Z hlediska metody odvozování je v tomto typu digitalizace obzvláště důležité, aby v žádné generaci souborů (mezi které patří původní snímky, ořezané snímky a snímky převedené do jiného formátu) nedošlo ke ztrátě klíčových vlastností, například převodem do ztrátové obrazové komprese. V případě, že budou soubory obrazové komponenty výsledného digitálního originálu uloženy v bezztrátové kompresi

(např. v bezztrátovém JP2), a přitom byly původní snímky uloženy ve ztrátovém formátu JPEG, lze konstatovat, že došlo ke ztrátě klíčových vlastností. Obrazová komponenta výsledného objektu CDO se však při validaci bude formálně jevit jako bezztrátová, ačkoliv reálně půjde o ztrátová data. To je jeden z příkladů toho, proč je důležitý popis metody odvozování v dokumentaci. Jen znalost kontextu a technologických podmínek digitalizace dále umožní porozumět například tomu, proč archivačním formátem starší digitalizace v prostředí českých knihoven byl ztrátový formát JPEG a transformace těchto starších dat do bezztrátového JP2 znamená nižší kvalitu než novější digitalizace, jejímž výstupem je bezztrátový JP2, který je i reálně bezztrátový.

Popis referenčního objektu, jehož součástí jsou klíčové vlastnosti a metoda odvozování, musí být nejen zaznamenán v dokumentaci, ale tato dokumentace musí být také dostupná čtenářům, aby na základě ní mohli digitalizát zobrazený v digitální knihovně posuzovat.

Dále je třeba uvést, že úplný referenční objekt logicky může existovat až po finalizaci balíčku SIP. V době produkce rastrové digitalizace je posuzování autenticity založeno na posuzování odvozování dílčích komponent objektu CDO (v případě obrazových dat jde o soubory několika generací, v případě textové a strukturální komponenty zpravidla jen jedné generace, neboť tyto se obvykle vytvářejí v jednom kroku). Do fáze vytváření digitálního originálu tedy spadají dílčí otázky jako: Jak věrně zachycuje původní snímek tištěnou předlohu? Byla při všech úpravách zachována bezztrátová komprese? Nebyly provedeny řezy, které odstraňují část stránky? Je strukturální komponenta vytvořena správně? Je chybovost procesu OCR vyšší, než by měla být?

## 4.2 Vyjádření identity digitalizátu knihy

Oddíl věnovaný referenčnímu objektu předpokládá jeho specifikaci před zahájením digitalizace a záznam v dokumentaci digitalizačního projektu – jedná se o typovou specifikaci (tj. platnou pro všechny digitalizáty projektu, resp. skupinu digitalizátů). Pro každý jednotlivý digitalizát knihy je však také potřeba zaznamenat a udržovat informace popisující jeho identitu. Tyto informace musejí být obsaženy v metadatech a uloženy v balíčku AIP v repozitáři; také však musejí být dostupné čtenářům, tj. být prezentované spolu se zobrazeným dokumentem v digitální knihovně ve formě metadat, resp. odkazu na dokumentaci digitalizačního projektu.

Vyjádření identity digitalizátu knihy musí logicky zahrnovat jak popis konkrétního referenčního objektu, tak typové vyjádření jeho vlastností, které sdílí s ostatními digitalizáty knih v rámci daného digitalizačního projektu.

Následující tabulka navrhuje, jaké minimální atributy identity digitalizátu knihy je nutné zaznamenat, aby bylo možné udržovat a posuzovat jeho autenticitu v dlouhodobém horizontu. Jinými slovy, aby bylo vůbec možné hovořit o specifickém dokumentu – bez dostatečného vyjádření identity nelze plně posuzovat autenticitu.

Minimální atributy pro vyjádření identity digitalizátu knihy			
1	Vyjádření atributů tištěné předlohy		
	1.1	Atributy exempláře knihy	
		1.1.1.	Identifikátor exempláře knihy
		1.1.2	Název vlastníka exempláře knihy
		1.1.3	Údaj o stavu exempláře
		1.1.4	Údaj o akvizici
	1.2	Atributy vydání knihy	
		1.2.1	Perzistentní identifikátor vydání knihy
		1.2.2	Údaj o barevných vlastnostech knihy
		1.2.3	Minimální bibliografický popis
	1.3	Údaj o časovém ukotvení tištěné předlohy	
2.	Vyjádření atributů digitálního originálu		
	2.1.	Perzistentní identifikátor digitalizátu knihy	
	2.2	Název vlastníka digitalizátu knihy	
	2.3	Popis klíčových vlastností digitalizátu knihy	
	2.4	Popis typové digitální identity digitálního originálu	
	2.5	Datum vzniku digitálního originálu	
	2.6	Název digitalizačního projektu	
3.	Specifikace věrnosti		
	3.1.	Popis úrovně abstrakce digitalizace podle FRBR	
	3.2	Popis metody odvozování	

Tab. 1

Minimálními požadavky z hlediska **atributů exempláře tištěné knihy** jsou identifikátor exempláře knihy, název vlastníka exempláře knihy, údaj o stavu exempláře a údaj o akvizici. Tyto údaje tvoří základ provenienčních informací o tištěné předloze jako o konkrétním fyzickém objektu, který byl zdrojem digitalizace. Údaj o stavu exempláře je rozhodný pro posuzování kvality digitálního převodu (v případě digitalizace na úrovni vyjádření zde může být informace o tom, že některé části textu nebyly čitelné), údaj o akvizici je v případě rastrové digitalizace důležitý také pro vysvětlení případů, kdy kniha (a její digitalizát)

obsahuje údaje o jiné knihovně, než která je udávána jako vlastník exempláře (např. obsahuje razítko jiné knihovny, neboť kniha byla získána z vyřazených knih).

Minimálními požadavky z hlediska **atributů vydání tištěné knihy** jsou perzistentní identifikátor vydání knihy, údaj o barevných vlastnostech knihy a minimální bibliografický popis (v případě digitalizace na úrovni exempláře je myšlen specifický exemplářový popis).

Perzistentní identifikátor je klíčový pro jednoznačné propojení digitalizátu knihy s jeho předlohou pro potřeby snadného vyhledávání čtenáři, celkovou koncepcí rozvoje fondu knihovny (jako nástroj, jak bezpečně zjistit, zda již nějaká kniha nebyla zdigitalizována) a pro celkové potřeby správy digitalizátů knih.

Minimální bibliografický popis slouží jako informační doplněk k tomuto identifikátoru, který je důležitý pro potřeby čtenáře. Údaj o barevných vlastnostech musí zahrnovat popis jakýchkoliv barevných vlastností knihy (tj. včetně různé barvy odlišných částí textu) a slouží pro posuzování věrnosti digitalizačního převodu, spolu s údajem o časovém hledisku tištěné předlohy (1.3). Je důležitý i v případě digitalizace na úrovni vyjádření (některé knihy obsahují barevný text pro odlišení nebo zvýraznění, což lze považovat za strukturní vlastnost textu).

Ačkoliv je běžnou praxí snímat knihy ve stavu v době digitalizace, je údaj o časovém ukotvení důležitým prvkem z dlouhodobého hlediska, kdy kontext digitalizace současné doby již nebude znám. Tento údaj musí být uveden v dokumentaci digitalizačního projektu (vztahuje se na všechny knihy daného digitalizačního projektu).

Z hlediska **vyjádření atributů identity digitálního originálu** je perzistentní identifikátor digitalizátu knihy nepostradatelný z hlediska dlouhodobé správy i otázek identity digitalizátu v rámci internetu. Tento perzistentní identifikátor musí splňovat obecné požadavky, které se očekávají od perzistentní identifikace digitálních dokumentů v internetovém prostředí (globální jedinečnost, možnost trvalého přesměrování užitím resolveru). Název vlastníka digitalizátu, údaj o datu vzniku digitálního originálu a název digitalizačního projektu jsou nezbytné provenienční informace o digitalizátu. Popis klíčových vlastností digitalizátu knihy je nezbytný z hlediska dlouhodobého uchování digitalizátu (jako popis toho, co má být zachováno napříč transformacemi digitálního originálu) i zpřístupňování čtenářů (jako popis toho, co mohou a mají čtenáři očekávat od digitalizátu prezentovaného v digitální knihovně). Popis typové digitální identity digitálního originálu se vztahuje na digitální identitu objektu CDO v balíčku SIP, která má být vytvořena v rámci digitalizace, přičemž musí být v souladu s klíčovými vlastnostmi. Objekt CDO bude v budoucnu migrován do jiného formátu, nicméně tyto údaje o původní digitální identitě se pak stanou součástí provenienčních informací, jako stopa po zaniklém objektu.

Uvedené atributy musejí být zapsány do metadat v balíčku SIP, s možnou výjimkou klíčových vlastností pro rastrovou digitalizaci (v případě užití formátu TEI jsou specifikovány přímo ve standardu, zatímco standard PREMIS užívaný pro rastrovou digitalizaci žádnou specifikaci klíčových vlastností neobsahuje). Popis typové digitální identity digitálního originálu a popis klíčových vlastností digitalizátu knihy však

musejí být obsaženy také v digitalizační dokumentaci, která musí být vytvořena ještě před zahájením digitalizace. Klíčové vlastnosti zaznamenané v dokumentaci v obecné rovině (pro rastrovou digitalizaci plná barevnost, dobrá čitelnost textu, původní posloupnost stran, úplnost stran, listovatelnost, plnotextová prohlédatelnost a kopírovatelnost textu) jsou řešení, jak nahradit v současnosti nedostupné podrobné metadatové schéma pro jejich zápis, a současně způsobem, jak čtenářské obci srozumitelným způsobem popsat, jaké vlastnosti lze od prezentovaného digitalizátu očekávat. V případě digitalizace na úrovni vyjádření s užitím formátu TEI by měly být v dokumentaci digitalizačního projektu zaznamenány ty strukturní vlastnosti textu, jejichž definice byly převzaty ze standardu TEI. Klíčové vlastnosti pak musejí být také součástí dokumentace archivu a musejí být odkazovány v digitální knihovně (odkazem na dokumentaci digitalizačního projektu).

**Specifikace věrnosti** zahrnuje popis zvolené úrovně abstrakce digitalizace podle FRBR. Popis metody transformace musí vycházet z údaje o časovém hledisku tištěné předlohy. Pokud je zvoleno snímání aktuálního stavu knihy, musí být metoda transformace taková, aby digitální originál co nejvěrněji zachovával současný vzhled knihy. Pokud je zvolena digitalizace do podoby předpokládaného stavu knihy v době vzniku, musí metoda transformace zahrnovat digitální restaurování a musí být uvedeno, jakým způsobem se zjistil tento předpokládaný stav, v případě digitalizace na úrovni vyjádření musí být uvedeno, zda text byl plně čitelný. Všechny tyto informace musejí být zaznamenány v digitalizační dokumentaci před zahájením digitalizace. Do metadat je vyjádřit lze jen obtížně a nebylo by to ani účelné.

### 4.3 Základní požadavky pro udržení autenticity

Ve fázi produkce (digitalizace) lze uvést následující požadavky. Při finalizaci balíčku SIP může dojít k tomu, že validace ukáže na chybu. V takovém případě je možné provést opravu, nebo zdigitalizovat knihu znovu. V případě masové rastrové digitalizace lze ze zkušenosti říci, že nejlepší volbou je většinou provedení opětovné digitalizace. Všechny opravy musejí být zaznamenány do metadat a takovýto záznam je složitější vytvořit, a ještě složitější jej interpretovat – jednodušší je provést novou digitalizaci.

Teprve pokud balíček SIP obsahuje validní objekt CDO, tedy digitální originál vytvořený v souladu s dokumentací digitalizačního projektu a mající všechny požadované klíčové vlastnosti, nastává (dlouhodobá) fáze udržování autenticity digitalizátu knihy. Autenticita musí být zachována v procesu dodávání balíčku SIP do repozitáře, při převodu do balíčku AIP v repozitáři i při vytvoření balíčku DIP (prezentační varianty) pro digitální knihovnu.

Základním doporučením je vytvářet objekt CDO v balíčku SIP v takovém formátu, který je archivačním formátem. Ve fázi příjmu balíčku SIP do repozitáře a jeho převodu do balíčku AIP pak nebude nutno provádět formátovou normalizaci, a sníží se tak riziko chyb. Takovýto postup také ušetří čas a technické, finanční a organizační zdroje repozitáře. Je vhodný také proto, že digitalizační projekt mívá obvykle lepší personální obsazení než repozitář. Optimálně by také digitalizační pracovníci měli provádět vizuální kontrolu digitalizátu s tištěnou knihou v ruce.



Doporučením pro fázi zpřístupňování je, aby digitální knihovna byla navržena jako součást repozitáře (tedy archivu OAIS), tj. aby plnila služby funkčního celku „Zpřístupnění“ normy ISO 14721 (ISO, 2012a, s. 60-61). To znamená, že balíčky DIP musejí být v digitální knihovně odvozovány přímo z balíčku AIP (a nikoliv z balíčku SIP). Pro českou praxi to znamená, že vytváření prezentačních variant by se mělo uskutečňovat nikoliv v digitalizaci, ale až v archivačním systému, který by měl být integrován s digitální knihovnou tak, aby oba systémy tvořily jeden celek.

Pro prezentaci v digitální knihovně je nutné, aby organizace pravidelně vybírala takový zpřístupňovací software, který je otestován na schopnost zobrazovat digitalizáty knihy v souladu s klíčovými vlastnostmi (takový test se neobejde bez lidské vizuální inspekce). Prezentace v digitální knihovně je výchozím bodem pro posuzování autenticity z hlediska čtenáře. Pokud digitální knihovna nabízí kromě prezentace digitalizátů knihy také možnost jejich stažení do zařízení čtenáře, musí současně poskytovat aktuální informace o vhodném softwaru, jehož pomocí si čtenář na svém zařízení může stažený digitalizát zobrazit, a také informaci o digitálním otisku daného dokumentu, na základě které lze ověřit bezchybnost stažení.

V budoucích fázích životního cyklu digitalizátu knihy (kdy dojde k zastarávání formátů nebo softwaru) budou muset být provedeny migrace takovým způsobem, aby byly zachovány klíčové vlastnosti, obsažené ve specifikaci referenčního objektu. Tyto operace musejí zahrnovat testování reprodukce v nových nástrojích a srovnání kvality reprodukce v novém a zastarávajícím softwaru, a to v podobě posouzení lidskými recipienty. Současně musí repozitář udržovat informace o všech těchto procesech odvozování: zaprvé ve formě metadat zaznamenávajících provenienci digitálního dokumentu (tj. popis transformací i technické informace popisující všechny předešlé generace objektu CDO, které již neexistují – takže po nich zůstanou jen tyto „metadatové stopy“), zadruhé ve formě dokumentace, která musí obsahovat informace o nastavení migračních procesů a postupech jejich testování, informace o vytváření prezentačních variant a důvody, proč k těmto procesům došlo.

Specificky pro české prostředí lze zdůraznit následující pravidla pro životní cyklus dokumentů s ohledem na udržení a posuzování autenticity: soustředit se na důkladnou katalogizaci (rekatalogizaci) tištěných knih před zahájením digitalizace, což může zahrnovat i zavedení celkové koncepce zkvalitnění záznamů v katalogu digitalizující knihovny; udržovat podrobnou digitalizační dokumentaci (včetně popisu všech postupů užitých při digitalizaci – nelze se spoléhat jen na automaticky generovaná metadata); sloučit archivační a prezentační systém do jednoho integrovaného systému archivu OAIS; řídit opravy digitalizátů knih prostřednictvím systému perzistentní identifikace digitálních dokumentů; využít další perzistentní identifikátor na úrovni vydání knihy. Poslední dvě doporučení rozvádí následující text.

Specifickým problémem, běžně řešeným v českém prostředí, jsou opravy digitálního originálu, které jsou prováděny ve fázi, kdy digitalizáty knihy již byly prezentovány čtenářům v digitální knihovně. Pokud jde o opravy, které znamenají významnou změnu dokumentu (např. dodání chybějících stran nebo rozdělení dokumentu do dvou a více nebo naopak spojení více dokumentů do jednoho), lze z hlediska zajištění

citační integrity a obecně důvěryhodnosti digitální knihovny jako vhodné řešení užít následující pravidlo: původní dokument musí být smazán z repozitáře (včetně digitální knihovny) a musí být vytvořen nový dokument (opravou nebo opětovnou digitalizací, kterou doporučujeme ve všech případech, kdy to možné je), kterému bude přidělen nový perzistentní identifikátor digitálního originálu se současným uvedením vazby na původní identifikátor. Dereference na základě identifikátoru původního dokumentu v resolveru pak bude zahrnovat pouze výpis metadat z registru identifikačního systému, které budou popisovat původní dokument, důvod jeho zániku (typ chyby / opravy) a odkaz na nový identifikátor. Teprve z nového identifikátoru bude možno provést přesměrování na URL adresu nového dokumentu. Tento mechanismus umožňuje a) zjistit, že citovaný dokument kdysi existoval („autenticita citace“), b) jasně upozornit čtenáře na skutečnost změny dokumentu, a tak snížit riziko uvedení čtenáře v omyl a vyhnout se narušení důvěryhodnosti citační praxe. Tento mechanismus je v praxi zaveden v pravidlech českého identifikačního systému ČIDLO, které je záhodno dodržovat (Vašek & Cubr, 2014).

Z hlediska perzistentní identifikace předlohy na úrovni vydání knihy (viz oddíl 4.2) lze za základní identifikátory považovat čísla ISBN (v českém prostředí je lze užít pro knihy od konce 80. let 20. století) a kódy NBN (lze užít pro všechny knihy zařazené do národní bibliografie – viz oddíl 3.6). V českém prostředí je však možné do katalogizačních záznamů zaznamenat ještě jeden perzistentní identifikátor – „československý kód“, který byl v druhé polovině 20. století přidělován českým (československým) knihám na úrovni titulu (vydání) a který je uveden v tiráži, případně i na přebalu. Systém přidělování tohoto kódu byl do jisté míry analogický k (mnohem později zavedenému!) systému ISBN: identifikátor se skládal ze tří částí tvořených čísly a oddělených pomlčkou. První část tvořilo dvojčífné číslo nakladatele, druhou část pořadové číslo knihy v daném nakladatelství a třetí dvojčíslí označovalo letopočet vydání (Pistorius, 2011, s. 57). Tento identifikátor sice nemá kontrolní součet, jako mají identifikátory ISBN, ale má stejnou výhodu – je uveden přímo v knize. Velmi překvapivě se dosud v české katalogizační praxi neužívá.

Doporučením pro české knihovny je tedy postupně doplňovat „československý kód“ do katalogizačních záznamů jako další perzistentní identifikátor vydání knihy (např. do pole 024 v MARC 21). Ačkoliv tento identifikátor nepokrývá celou českou knižní produkci, je výjimečný právě v tom, že na rozdíl od čísla ČNB (což je virtuální identifikátor, obsažený pouze v bibliografickém záznamu) je „československý kód“ možné opsat přímo z knihy, čím je zaručeno jednoznačné propojení záznamu a knihy.

## 5 Závěr

Cílem našeho článku bylo navrhnout základní konceptuální rámec pro udržování a posuzování autenticity digitalizátů knih v celém jejich životním cyklu, s ohledem na dlouhodobý horizont tohoto cyklu (návrh se vztahuje na ty digitalizáty knih, které jsou určeny k dlouhodobému uchovávání v repozitáři, nikoliv na dočasně uchovávané dokumenty, které budou časem smazány). Návrh konceptuálního rámce je založen na premise, že pro konkrétní typ dokumentu je nutno vypracovat specifickou konceptualizaci autenticity, a z návrhu je patrné, že tento přístup je správný, protože některé prvky návrhu nelze vztahovat na jiné typy dokumentů.

Předložený návrh je částečně potvrzením konceptu sociální konstrukce autenticity MacNeilové a Makové, podle něhož může legitimně existovat několik odlišných přístupů k autenticitě téhož předmětu v závislosti na teorii nebo užití perspektivě. Zásadní rozdíl mezi předkládaným rámcem a pojetím těchto autorek je však v tom, že jejich koncept se vztahuje na již existující objekty, zatímco navrhovaný rámec je předpisem pro samotný způsob vytváření objektů takovým způsobem, aby procesy udržování a posuzování autenticity byly co nejvíce transparentní, resp. aby byly vůbec proveditelné.

Základem návrhu konceptuálního rámce je užití modelu OAIS pro popis životního cyklu dokumentu, a především kritická analýza stávajících tematizací autenticity ve vybraných odborných zdrojích s cílem zachytit všechny možné aspekty, které by se mohly týkat autenticity digitalizátů knih, a to ve všech fázích jejich životního cyklu. Výstupem analýzy je redefinice pojmu autenticity ve vztahu k digitálnímu dokumentu obecně a návrh konceptuálního rámce pro autenticitu zaměřeného specificky na autenticitu digitalizátů knih. Návrh rámce ukazuje na potřebu již před zahájením digitalizace precizně stanovit základní rámec pro otázky autenticity, a to v podobě specifikace „referenčního objektu“ autenticity (obecná rovina) a specifikace atributů identity dokumentu (rovina jednotliviny). Specifikace referenčního objektu se přitom bude nutně lišit podle zvoleného přístupu k věrnosti digitalizace, který je determinován zvolenou úrovní abstrakce podle FRBR a metodou odvozování a v souladu s tím stanovených klíčových vlastností, jejichž zachování při dlouhodobém uchovávání v repozitáři i trvalém zpřístupňování čtenářům v digitální knihovně ustanovuje závazkem spravující organizace.

Rámec pro otázky autenticity digitalizátu knihy by tak v optimálním případě měl být částečným konstruktem (s nutnou predeterminací předlohou a možnostmi digitálních technologií) organizace řídící celý životní cyklus digitálního dokumentu, který znamená precizní a transparentní specifikaci toho, jaký z několika legitimních způsobů digitalizace (podle úrovně abstrakce FRBR) byl zvolen. Tato specifikace musí být navržena a zdokumentována před samotným zahájením digitalizace.

## 6 Literatura

- Adobe Systems Incorporated. (2005). *Adobe® RGB (1998) Color Image Encoding* [online]. Dostupný z: <https://www.adobe.com/digitalimag/pdfs/AdobeRGB1998.pdf> [cit 2018-06-27].
- Bearman, D. (1999). Reality and Chimeras in the Preservation of Electronic Records [online]. *D-Lib magazine*, 5(4). Dostupný z: <http://www.dlib.org/dlib/june98/06bearman.html> [cit 2018-06-27].
- Bearman, D., & Trant, J. (1998). Authenticity of Digital Resources : Towards a Statement of Requirements in the Research Process [online]. *D-Lib magazine*, 4(6). Dostupný z: <http://www.dlib.org/dlib/june98/06bearman.html> [cit 2018-06-27].
- Benjamin, W. (1979). *Dílo a jeho zdroj*. Praha: Odeon.
- British Library. (2015). *PDF format preservation assessments* [online]. Dostupný z: [http://wiki.dpconline.org/images/e/e8/PDF\\_Assessment\\_v1.3.pdf](http://wiki.dpconline.org/images/e/e8/PDF_Assessment_v1.3.pdf) [cit 2018-06-27].
- Brown, A. (2006). *The PRONOM PUID Scheme: a scheme of persistent unique identifiers for representation information* [online]. Dostupný z: [http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom\\_unique\\_identifier\\_scheme.pdf](http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf) [cit 2018-06-27].
- Caplan, P. (2008). The preservation of digital materials. *Library Technology Reports*, 44(2), 4-38.
- Cubr, L. (2010). Budování důvěryhodného systému trvalé identifikace digitálních dokumentů. *Knihovna*, 21(1), 23-31.
- Data Seal of Approval Board. (2016). *Data Seal of Approval : Guidelines version 2017-2019* [online]. Dostupný z: [https://assessment.datasealofapproval.org/guidelines\\_54/pdf/](https://assessment.datasealofapproval.org/guidelines_54/pdf/) [cit 2018-06-27].
- Deegan, M., & Tanner, S. (2006). Key issues in digital preservation. In M. Deegan, & S. Tanner (Eds.), *Digital preservation* (s. 1-31). London: Facet.
- Digital Library Federation Benchmark Working Group. (2002). *Benchmark for Faithful Digital Reproductions of Monographs and Serials* [online]. Dostupný z: <http://old.diglib.org/standards/bmarkfin.pdf> [cit 2018-06-27].
- Digital Preservation Handbook*. (2018) [online]. Dostupný z: <http://www.dpconline.org/handbook/> [cit 2018-06-27].
- Duranti, L. (1995). Reliability and Authenticity: The Concepts and Their Implications. *Archivaria* 39, 5-10.
- Duranti, L. (1996). The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project. *Archivaria* 42, 46-67.

Duranti, L. (2001). The impact of digital technology on archival science. *Archival Science*, 1(1), 39-55.

Duranti, L., & Thibodeau, K. (2006). The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES. *Archival Science*, 6(1), 13-68.

Federal Agencies Digitization Initiative. (2010). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* [online]. Dostupný z: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf) [cit 2018-06-27].

Fernie, K. (Ed.) (2008). *Technical Guidelines for Digital Cultural Content Creation Programmes* [online]. Dostupný z: <http://www.minervaeurope.org/publications/MINERVA%20TG%202.0.pdf> [cit 2018-06-27].

Giaretta, D. (2011). *Advanced Digital Preservation*. New York: Springer.

Gladney, H. M. (2007). *Preserving digital information*. Berlin: Springer.

Grycz, C. J. (2006). Digitising Rare Books and Manuscripts. In L. MacDonald (Ed.), *Digital heritage: applying digital imaging to cultural heritage* (s. 33-68). Amsterdam: Elsevier.

Hakala, J. (2001). *Using National Bibliography Numbers as URNs* [online]. Dostupný z: <http://www.ietf.org/rfc/rfc3188.txt> [cit 2018-06-27].

IFLA. (2009). *Functional Requirements for Bibliographic Records: final report* [online]. Dostupný z: [https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf) [cit 2018-06-27].

InterPARES project. (2005a). *Authenticity Task Force Report* [online]. Dostupný z: [http://www.interpares.org/book/interpares\\_book\\_d\\_part1.pdf](http://www.interpares.org/book/interpares_book_d_part1.pdf) [cit 2018-06-27].

InterPARES project. (2005b). *Preservation Task Force report* [online]. Dostupný z: [http://www.interpares.org/book/interpares\\_book\\_f\\_part3.pdf](http://www.interpares.org/book/interpares_book_f_part3.pdf) [cit 2018-06-27].

InterPARES project. (2005c). *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* [online]. Dostupný z: [http://www.interpares.org/book/interpares\\_book\\_k\\_app02.pdf](http://www.interpares.org/book/interpares_book_k_app02.pdf) [cit 2018-06-27].

InterPARES project. (2005d). *The InterPARES Glossary* [online]. Dostupný z: [http://www.interpares.org/book/interpares\\_book\\_q\\_gloss.pdf](http://www.interpares.org/book/interpares_book_q_gloss.pdf) [cit 2018-06-27].

ISO. (2003). *ISO 14721:2003: Space data and information transfer systems - Open archival information system (OAIS) - Reference model*. Geneva: ISO.

ISO. (2012a). *ISO 14721:2012: Space data and information transfer systems - Open archival information system (OAIS) - Reference model*. Geneva: ISO.

- ISO. (2012b). *ISO 16363:2012: Space data and information transfer systems - Audit and certification of trustworthy digital repositories*. Geneva: ISO.
- Kraus, J., & et al. (2005). *Nový akademický slovník cizích slov A-Ž*. Praha: Academia.
- Kroupová, L., & et al. (2005). *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. Praha: Academia.
- MacNeil, H., & Mak, B. (2007). Constructions of Authenticity. *Library Trends*, 56(1), 26-52.
- Malá československá encyklopedie. (1984-1987). Praha: Academia.
- Oxford English Dictionary (2018) [Website]. Dostupný z <http://www.oed.com/> [cit 2018-06-27].
- Ministerstvo kultury. (2015). *Věřejné informační služby knihoven (VISK): podprogram č. 7: národní program ochrany a digitalizace dokumentů ohrožených degradací kyselého papíru – KRAMERIUS* [online]. Dostupný z: <http://visk.nkp.cz/dokumenty/visk7/2016/VISK7-podm2016.doc> [cit 2018-06-27].
- NISO. (2007). *A framework of guidance for building good digital collections: a NISO recommended practice*. Dostupný z <http://www.niso.org/publications/rp/framework3.pdf> [cit 2018-03-15].
- Pistorius, V. (2011). *Jak se dělá kniha: příručka pro nakladatele*. Příbram: Pistorius & Olšanská.
- PREMIS Editorial Committee. (2015). *PREMIS Data Dictionary for Preservation Metadata* [online]. Dostupný z: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf> [cit 2018-06-27].
- Rothenberg, J. (1999). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources* [online]. Dostupný z: <https://www.clir.org/wp-content/uploads/sites/6/pub77.pdf> [cit 2018-06-27].
- Rothenberg, J. (2000). Preserving Authentic Digital Information. In *Authenticity in a digital environment* (s. 51-68). Washington, D.C.: Council on Library and Information Resources.
- UNESCO. (2016). *Guidelines for the selection of digital heritage for longterm preservation* [online]. Dostupný z: [https://www.unesco.nl/sites/default/files/uploads/Comm\\_Info/persistcontentguidelinesfinal1march2016.pdf](https://www.unesco.nl/sites/default/files/uploads/Comm_Info/persistcontentguidelinesfinal1march2016.pdf) [cit 2018-06-27].
- Vácha, J., & et al. (1971). *Slovník spisovného jazyka českého*. Praha: Academia.
- Vašek, Z., & Cubr, L. (2015). *Metodika pro přidělování a správu životního cyklu unikátních perzistentních identifikátorů digitálních dokumentů podle standardu URN:NBN* [online]. Dostupný z: <https://resolver.nkp.cz/urn:nbn:cz:nk-0027gj> [cit 2018-06-27].

Vychodil, B. (2013). *Produkce digitálních obrazových dat a jejich kontrola*. Praha: Univerzita Karlova v Praze. Disertační práce.