

Projekt ArcLib – budování systému pro dlouhodobou archivaci digitálních dat v českých knihovnách

Mgr. Marek Melichar

marek.melichar@ruk.cuni.cz

Mgr. Jan Hutař

jan.hutar@dia.govt.nz

RNDr. Michal Růžička

mruzicka@mail.muni.cz

Bc. Zdeněk Hruška

Zdenek.Hruska@mzk.cz

Mgr. Zdeněk Vašek

zdenek.vasek@nkp.cz

RNDr. Miroslav Bartošek, CSc.

bartosek@ics.muni.cz

Ing. Martin Lhoták

lhotak@knav.cz

Článek informuje o záměrech projektu ArcLib financovaném z programu aplikovaného výzkumu Ministerstva kultury ČR NAKI II. Nejprve stručně shrnujeme předcházející aktivity v oblasti dlouhodobé archivace digitálních dat v českých knihovnách. Jsou popsány projekty, na které ArcLib navazuje. Stručně jsou popsány plánované cíle projektu ArcLib a předpokládaný způsob řešení. Cílem projektu ArcLib je mimo jiné vytvoření open source řešení pro zajištění logické i bitové dlouhodobé ochrany digitálních dokumentů v souladu s potřebami všech typů knihoven v ČR.

Úvod

Cílem tohoto textu je informovat o záměrech projektu několika českých knihoven, který má vytvořit volně dostupné řešení pro zajištění logické i bitové ochrany digitálních dat. Projekt ArcLib navazuje na předcházející aktivity českých knihoven, o některých zde podrobněji pojednáme, a reaguje na potřeby knihoven při správě stále rostoucích sbírek dnes již poměrně různorodých digitálních dat.

Situace českých knihoven se za posledních deset let významně proměnila. Od roku 2006, kdy vznikl koncept Národní digitální knihovny (STOKLASOVÁ, 2006), došlo k několika násobnému navýšení objemu digitálních dat, které knihovny spravují. Díky projektům financovým z fondů EU, Norských fondů a z dalších zdrojů byly schopny i krajské a menší instituce ve větší míře pořídit vlastní digitalizační kapacity. Především akademické knihovny začaly systematicky pracovat s výhradně digitálními dokumenty. Další knihovny se soustředily na speciální dokumenty, jako jsou mapy nebo zvukové dokumenty. Vzrůstající objem digitálních dat nutí knihovny uvažovat nad způsobem jejich dlouhodobé správy. Projekt ArcLib by měl dát knihovnám do rukou nástroj, který jim dlouhodobou ochranu a správu digitálních dat usnadní.

Dlouhodobá archivace digitálních dat v českých knihovnách

Projekt ArcLib navazuje na řadu aktivit českých knihoven z posledních patnácti let. Motivace k tomu, zabývat se vážně dlouhodobou archivací, přišla do českých knihoven především s digitalizovanými daty. Systematičtější digitalizace začala kolem roku 2000. Ve stejné době Národní knihovna České republiky (NK ČR) také zahájila archivaci webu. Digitalizace cenných historických dokumentů v několika projektech probíhala již dříve (více v HUTAŘ, MELICHAR, 2015; MELICHAR, HUTAŘ, 2013; MELICHAR, HUTAŘ, 2014). Hlavním programem na podporu digitalizace novodobých fondů byl Kramerius, zaměřený především na starší tituly z 19. století a nejvíce ohrožené dokumenty na kyselém papíře. Zpřístupnění a uložení dat vznikajících v programu Kramerius řešila NK ČR. Jednotlivé knihovny začaly postupně provozovat vlastní digitální knihovny a některé začaly skenovaná data také archivovat. Národní knihovna kolem roku 2008 dospěla k projektu Národní digitální knihovna (NDK), který měl spojit digitalizaci, dlouhodobou archivaci stávajících i nových dat a jejich odpovídající zpřístupnění.

Koncept NDK z roku 2006 předpokládal vybudování centrálního důvěryhodného digitálního repozitáře pro zajištění trvalého uchování jádra kulturního dědictví v digitální podobě a vedle toho existenci řady „dalších digitálních dokumentů oborového, regionálního, institucionálního i jiného charakteru“ (STOKLASOVÁ, 2006) za něž budou odpovědné další organizace. Od roku 2011 probíhala vlastní realizace projektu Vytvoření NDK. V jeho rámci vznikla zejména dvě digitalizační pracoviště v NK ČR v Praze a v Moravské zemské knihovně v Brně (MZK) a také systém na dlouhodobou ochranu digitálních dat (dále LTP systém), ve kterém jsou uloženy a dlouhodobě uchovávány výsledky projektu, což v současnosti představuje takřka 35 mil. stran. Data jsou uložena na magnetických páskách a spravována pomocí systému SAFE LTP vytvořeného v rámci projektu. V projektu vytvořená data jsou uložena ve třech kopiích, ve dvou v Praze, v jedné v Brně. Vedle dat z projektu Vytvoření NDK jsou do LTP systému ukládána také digitalizovaná data dalších knihoven vzniklá v programu VISK 7 (Kramerius). Cílem NK ČR je provést certifikaci LTP systému podle normy ČSN ISO 16363, k čemuž směřuje postupnou přípravou veřejně dostupné dokumentace.

Akademické knihovny v České republice spravují také poměrně velké sbírky digitálních dat. Vedle vysokoškolských kvalifikačních prací ukládají a zpřístupňují výsledky publikační činnosti svých akademických pracovníků, výuková data a v budoucnu dojde jistě i na vědecká data (tedy výstupy z výzkumu). Téměř každá veřejná vysoká škola provozuje jeden nebo hned několik digitálních repozitářů. Některé z projektů akademicky orientovaných knihoven se věnují dlouhodobé archivaci svých dat. Například projekt Národního úložiště šedé literatury¹ pravidelně provádí audit pomocí metodiky DRAMBORA² a také zveřejnil metodiku týkající se mimo jiné i dlouhodobé archivace dat šedé literatury (PEJŠOVÁ, 2011). Rovněž komunita provozovatelů otevřených repozitářů se zajímá o to, jak implementovat požadavky referenčního modelu OAIS (ČSN ISO 14721). Zatím dva repozitáře na Univerzitě Karlově vyjádřily svůj závazek dlouhodobě ukládat svěřená data pomocí certifikace podle Data Seal of Approval (DSA)³. Partnerem projektu ArcLib je také Masarykova univerzita a její účast by měla zajistit, že ArcLib zohlední potřeby repozitářů akademických knihoven.

Projekt ArcLib přímo navazuje na aktivity projektu Česká digitální knihovna (CDK)⁴. Původní koncept NDK z roku 2006 předpokládal, že kolem jádra knihovního kulturního dědictví archivovaného v důvěryhodném repozitáři NK ČR, budou existovat další dokumenty, které budou ukládány a zpřístupňovány v takzvané České digitální knihovně. Tento název pak použila Knihovna akademie věd (KNAV) k označení projektu financovaného z prostředků programu NAKI Ministerstva kultury v letech 2012 – 2015. V rámci projektu CDK byl dále rozvíjen systém Kramerius, který byl již před tímto projektem dominantním nástrojem pro zpřístupňování skenovaných digitálních dat v českých knihovnách. V současnosti jej používá více než 40 knihoven. Dále byly v projektu CDK vyvíjeny nástroje pro podporu procesu digitalizace – produkční systém ProArc a systém RFlow pro řízení digitalizačního procesu. Projekt ArcLib by měl navázat na CDK přidáním samostatného archivačního řešení, které bude dostupné jako open source nejen institucím používajícím ProArc nebo Kramerius, ale i dalším institucím.

Další motivací pro projekt ArcLib byla práce na Konceptu rozvoje knihoven pro roky 2011 – 2015⁵. V oblasti dlouhodobé archivace vznikl požadavek na testování volně dostupných řešení a ten našel vyjádření v projektu LTP Pilot. Projekt LTP Pilot byl financován z Fondu rozvoje CESNET, byl řešen Masarykovou univerzitou ve spolupráci s Moravskou zemskou knihovnou a Jihočeskou univerzitou a trval 14 měsíců v letech 2014 – 2015. Jeho cílem byla „pilotní implementace a ověření nízkobariérového řešení pro dlouhodobou archivaci digitálních dat na bázi systému Archivematica“⁶ včetně průzkumu možností integrace tohoto systému s infrastrukturou digitálních úložišť CESNET⁷. Kromě testování konkrétního open-source řešení a vytvoření znalostní báze k němu, měl projekt také motivovat české knihovny ke zvyšování úrovně ochrany jimi spravovaných dat. Praktické zkušenosti s provozem několika demo instalací a s testováním systému Archivematica s několika sbírkami, při simulaci reálného provozu, budou základem pro další rozhodování zda a jak zapojit systém Archivematica do architektury software vznikajícího v projektu ArcLib.

Základní informace o projektu ArcLib

Hlavním řešitelem projektu ArcLib je Knihovna AV ČR, v. v. i. (KNAV) a dalšími partnery jsou Masarykova univerzita, NK ČR a MZK. Projekt je financován z programu aplikovaného výzkumu Ministerstva kultury NAKI II. Je naplánován na 5 let, na roky 2016 – 2020. Jeho průběh je rozdělen do dvou etap, v první etapě (2016 – 2018) budou probíhat práce na analýze stávajících standardů dlouhodobé archivace a systémů k integraci do ArcLib; příprava metodik; bude zahájen vývoj softwaru a testování. Ve druhé etapě (2019 – 2020) pak bude pokračovat vývoj, testování, produkční implementace a poté nasazení archivačního řešení ArcLib v poloprovozu.

Projekt ArcLib má čtyři hlavní cíle:

- Vývoj vlastního komplexního LTP (Long Term Preservation) open-source řešení ArcLib pro zajištění jak logické tak i vysoké úrovně bitové ochrany digitálních dat.
- Vytvoření metodiky pro dlouhodobou logickou ochranu digitálních dat pro české prostředí s ohledem na mezinárodní standardy (zejména referenční model OAIS – ČSN ISO 14721 a ČSN ISO 16363).
- Vytvoření metodiky a návrh řešení pro fyzické ukládání velkého množství dat a zajištění bit-level ochrany pro potřeby dlouhodobé ochrany.
- Ověření funkčnosti celého řešení v praxi formou poloprovozu minimálně v jedné ze zapojených institucí.

Jednotlivé zúčastněné instituce mají v projektu specifické role – KNAV je odpovědná za řízení projektu a koordinaci vývojových prací, Masarykova univerzita bude spolupracovat na návrhu softwarového řešení pro logickou ochranu digitálních dat, ale především bude odpovědná za návrh řešení a metodiky v oblasti bitové ochrany digitálních dat. Bude také pracovat na integraci vznikajícího řešení ArcLib se systémy repozitářů akademických knihoven (především systém DSpace). MZK a NK ČR budou spolupracovat na vzniku softwarového řešení a vznikajících metodikách. NK ČR bude také spolupracovat na zajištění interoperability se standardy NDK tak, aby bylo možné balíčky ArcLib vkládat v případě potřeby do LTP systému NDK a případně i naopak. Řešení ArcLib předpokládá integraci s dalšími již existujícími systémy, jako jsou Kramerius, ProArc, systémy NDK nebo systémy akademických knihoven, jako jsou digitální repozitáře na bázi DSpace. Integrace bude pro úspěch systému ArcLib klíčová.

¹ <http://www.nusl.cz/>

² <http://www.repositoryaudit.eu/>

³ <http://www.datasealofapproval.org/en/assessment/>, <http://dsa.cuni.cz/DSA-12.html?look=new>

⁴ <http://www.czechdigitallibrary.cz/softwareva-reseni/>

⁵ <http://www.ukr.knihovna.cz/koncepce-rozvoje-knihoven-cr-na-leta-2011-2015-/>

⁶ <http://ltp-portal.mzk.cz/ltp-pilot>

⁷ <https://du.cesnet.cz/cs/start>

Základní vstupní požadavky na vyvíjený systém budou definovány zejména nově vytvořenými metodikami pro logickou i bitovou ochranu dat, které budou vycházet především z relevantních ISO standardů.

Jedním z výstupů projektu ArcLib bude vytvoření metodiky pro dlouhodobou logickou ochranu metodiky pro zajištění ochrany bitů. Předpokládá se, že metodika pro dlouhodobou logickou ochranu digitálních dat bude mít tři části. V úvodní teoretické části budou vysvětleny základní koncepty dlouhodobé ochrany digitálních dat a související terminologie; bude popsán koncept informačního balíčku a existující příklady strategií dlouhodobé ochrany digitálních dat. Tato část metodiky by měla poskytnout oporu institucím při plánování a budování systému pro dlouhodobou archivaci digitálních dat.

Druhá část metodiky bude vázána na produkt vznikající v projektu ArcLib. Měla by popsat způsob převedení požadavků odvozených z referenčního modelu OAIS a jeho popisu informačního modelu do konkrétního informačního systému ArcLib a standardů informačních balíčků v ArcLib. Také by měla obsahovat návod pro zjišťování kvality pomocí self-auditů a doporučení pro vedení procesní dokumentace OAIS archivu.

Poslední, implementační část metodiky pak bude obsahovat konkrétní plány ochrany pro typy dat, která budou v systému spravována během poloprovozu. Doporučení by měla obsahovat už konkrétní postupy validace formátů, rozsah technických metadat, ale i doporučení pro personální a organizační zajištění provozu.

Dalším z výstupů projektu ArcLib bude metodika pro fyzické ukládání dat a zajištění jejich ochrany na bitové úrovni. Na rozdíl od logické ochrany zde není cílem „rozumět“ uchovávaným datům, ale zajistit bezpečné a neměnné uložení binárních streamů, kterými jsou uchovávaná data reprezentována. A to i při velkém objemu dat. Na základě analýzy rizik ohrožujících fyzické uložení dat budou řešeny jak technické aspekty, tj. doporučené úložné technologie a způsob jejich napojení na informační systém, tak administrativní aspekty, tj. výběr lokalit pro fyzické umístění úložišť, politiky kontroly integrity dat a plány zotavení dat při jejich poškození. Stejně jako v případě metodiky pro logickou ochranu budou zahrnuty i konkrétní postupy a technická řešení pro potřeby poloprovozu informačního systému ArcLib.

Možnosti řešení ArcLib

Projekt ArcLib se inspiroval některými projekty vývoje systémů pro dlouhodobou ochranu digitálních dat, které k tomuto účelu také využívají open-source software. Například v Zuse Institute Berlín probíhá v současné době budování a testování LTP systému, který je založený na open-source nástrojích jako je Archivematica⁸, iRODS⁹ a Islandora¹⁰. Portugalský Národní digitální archiv pro změnu využívá systém RODA¹¹, který je v této zemi vyvíjen již od roku 2006. Dalším zajímavým příkladem vývoje LTP systému je finská státní nezisková společnost CSC (IT Center for Science)¹², kde pro dlouhodobou archivaci využívají kombinaci open-source a in-house vyvíjeného softwaru¹³. Bohužel ani RODA ani Archivematica nejsou ideální. Po testování v projektu LTP Pilot víme, že Archivematica není v současnosti zralé řešení se všemi požadovanými funkcemi, které bychom mohli s klidným srdcem vložit do jádra architektury ArcLibu s očekáváním, že bude smysluplně použitelné i za 5 a více let. Ani RODA neobsahuje všechny moduly (chybí modul pro plánování uchovávání, která byl řešen v původní architektuře RODA dalším systémem CRIB, viz RAMALHO, 2008), navíc její uživatelská komunita není příliš aktivní. V obou případech jde o firmami vyvíjený open-source software, za nímž nestojí mnoho vývojářů nebo silných institucí. Použití těchto systémů s sebou nese rizika, která musíme při návrhu řešení pro ArcLib zvažovat. Další možností je vyjít při návrhu řešení pro ArcLib ze systému ProArc, který podobně jako digitální knihovna Kramerius využívá repositář Fedora¹⁴. Fedora je middleware; pochopitelně nemá všechny funkce pro zajištění dlouhodobé archivace. Tento systém ale vyvíjí mnohem větší a aktivnější komunita a při jeho využití by se dalo počítat s menším rizikem. Případná modifikace ProArcu pro použití v LTP systému bude vyžadovat podrobnější analýzu požadovaných funkcí a rozsáhlejší vývoj.

Architektura řešení ArcLib by měla být otevřená a modulární, tak aby bylo možné jednotlivé komponenty (databáze, index, funkční moduly, mikroslužby, BPM/workflow, nástroje pro zpracování formátů apod.) časem vyměnit bez ohrožení funkcí zbytku systému. Ideálním řešením by mohl být systém volně, přes aplikační rozhraní provázaných modulů spojených dohromady nějakou logikou nebo systém skládající se z mikroslužeb a nějakého rámce kolem nich.

Návrh řešení ArcLib bude vycházet z požadavků mapovaných na funkční komponenty referenčního modelu ČSN ISO 14721 (OAIS) a na jeho informační model. Rozpracováním OAIS vzniklo v minulosti několik metadatových standardů, které mají usnadnit implementaci některých aspektů digitálního archivu, jak jej definuje OAIS: jsou to např. na XML postavená metadatová schémata METS¹⁵ a PREMIS¹⁶. Podobně bude muset řešení ArcLib respektovat národní standardy knihoven v ČR pro produkci digitalizovaných dat – tzv. NDK standard¹⁷. Systém pro dlouhodobou správu digitálních dat se musí zaměřit především na moduly OAIS Ingest (vstup), Data management (správa dat) a Archival storage (archivní uložení). Jelikož OAIS standard popisuje funkce archivu obecně a nezávisle na konkrétním technickém řešení, bude třeba při návrhu řešení ArcLib podrobně definovat konkrétní požadavky právě v těchto oblastech. Kromě toho bude třeba rozhodnout, které události při správě uložených dat bude systém zaznamenávat a jak, které změny budou důvodem k vytvoření nové verze AIP balíčku (ak-

⁸ <https://www.archivematica.org/>

⁹ <http://irods.org/>

¹⁰ <http://islandora.ca/>

¹¹ <http://www.roda-community.org/>

¹² <https://www.csc.fi/>

¹³ <https://stanford.app.box.com/s/a13qvgbbuhwjmza5dfm6kaf3qg2y71z>

¹⁴ <http://fedorarepository.org/>

¹⁵ <http://www.loc.gov/standards/mets/>

¹⁶ <http://www.loc.gov/standards/premis/>

¹⁷ <http://www.ndk.cz/standardy-digitalizace>

tualizace bibliografických, kontextových, vysvětlujících metadat, revalidace formátů a nová extrakce technických metadat), a které události bude třeba realizovat vně archivačního systému (plánování uchovávání, sledování uživatelské komunity, migrace formátů streamu a plánování migrace).

Řešení vytvořené v projektu ArcLib by mělo knihovnám pomoci řídit zavedení a provoz systému pro dlouhodobou archivaci digitálních dat jako soubor kontrolovaných a dokumentovaných procesů. Proto by uživatelé ArcLib měli mít možnost využít v projektu vytvořené doporučené postupy pro zlepšování kvality procesů dlouhodobé archivace. Dokumentace a dodržování postupů jim může poté pomoci směřovat k auditu podle DSA nebo dokonce ČSN ISO 16363. Metodickou asistenci a návod pro vytvoření dokumentace, plánů a politik by jim vedle relevantních ISO norem měly dát i metodiky vznikající v projektu ArcLib, které budou zohledňovat i související požadavky týkající se informační bezpečnosti a řízení rizik provozu informačního systému.

Z hlediska bitové ochrany uložená data ohrožuje nejen selhání hardware nebo jeho zničení např. v důsledku živelné pohromy. Ke ztrátě dat velmi často vedou chyby obsluhy, nebo úmyslné útoky operátora nebo jiného subjektu. V rámci metodiky bitové ochrany bude proto nutné stanovit vhodné postupy pro minimalizaci škod způsobných působením takových událostí.

Vhodné se jeví ukládání identických kopií dat ve více geograficky oddělených lokalitách na různých typech úložišť spravovaných různými skupinami osob a s minimálně jednou off-line kopií. To vše doprovázené pravidelnými kontrolami dostupnosti a integrity dat.

Pro menší instituci s omezeným počtem pracovníků může být problematické vybudování a údržba více geograficky a administrativně oddělených datových úložišť. V tomto ohledu může být vhodná spolupráce více institucí např. na bázi modelu Private LOCKSS Network¹⁸. Další zajímavou alternativou může být vybudování dalšího úložiště pro institucionální data s využitím veřejných cloudových úložných služeb, jako je např. Amazon Simple Storage Service (S3)¹⁹ nebo Oracle Storage Cloud²⁰, případně Datových úložišť CESNET²¹.

V rámci projektu ArcLib se uvažuje o vybudování minimálně jednoho úložiště přímo v rámci instituce KNAV. Při budování lokálního úložiště dnes existuje několik široce používaných technologií, zejména pevné disky a různé typy magnetických pásek. Nabídka jiných úložných technologií (např. velkokapacitní knihovny optických disků, svícení dat na filmový pás apod.) je omezená a daná úložiště mají často velmi specifické provozní vlastnosti. Podstatné je také zvážit, jaký typ úložiště se vyplatí z hlediska provozních nákladů: Pevné disky spotřebovávají elektrickou energii po celou dobu provozu, magnetické pásky napájení nepotřebují, ale pořizovací cena páskových mechanik může být velmi vysoká.

Z hlediska nároků na údržbu lokálního úložiště je nutné zvážit, zda je lepší pořídit hotové úložné řešení (hardwarové diskové pole, páskovou knihovnu, ...) včetně specifického obslužného software dodavatele a platit jeho technickou podporu, nebo úložiště postavit na levných „hloupých“ komponentách obsluhovaných „inteligentním“ open-source software spravovaným přímo provozovatelem úložiště. Toto řešení je možné zejména u diskových úložišť, kde jsou navíc nyní k dispozici pokročilé open-source souborové systémy Btrfs²² a ZFS²³ cílené na budování velkokapacitních (exabyty nebo i více) úložišť s důrazem na bezpečnost a konzistenci uložení dat (průběžně kontroly všech dat pomocí automaticky počítaných kontrolních součtů, samoléčení souborového systému, tj. automatická obnova poškozených dat z kopií na jiných discích).

V projektu ArcLib plánujeme v maximální míře stavět na existujících systémech s dobrou uživatelskou základnou, a tedy dobrou prognózou životaschopnosti do budoucna. V tomto ohledu se jeví jako velmi zajímavý projekt Ceph²⁴. Ceph je již nyní velmi populární v rámci cloudové platformy OpenStack²⁵, která má dobré vyhlídky na velkou popularitu a značné rozšíření do budoucna, přitom je na ní nezávislý a dá se nasadit samostatně.

Ceph poskytuje plně distribuované síťové úložiště. Pomocí Ceph je možné spojit geograficky oddělené úložné uzly do jednoho úložiště s definovatelnou politikou replikace dat mezi úložnými uzly. Integrita dat je chráněna kontrolními součty, Ceph obsahuje nástroje pro jejich pravidelnou automatickou kontrolu. Ceph přitom poskytuje úložiště nejen v podobě souborového systému, ale obsahuje také funkce objektového úložiště, což může být velmi vhodné pro ukládání balíčků AIP v rámci LTP systému. Navíc je pro objektové úložiště Ceph k dispozici aplikační rozhraní kompatibilní s Amazon S3. LTP systém by pak byl schopen jednou implementací komunikačního rozhraní obsluhovat úložiště postavené nejen na Ceph, ale i na Amazon S3.

Nevýhodou Ceph je, že počítá s nasazením nad pevnými disky. Neumožňuje tak integraci např. páskových knihoven. Pokud bychom chtěli vytvořit distribuované úložiště obsahující také páskové knihovny (nebo jiná „exotická“ úložiště), tak se jako jediná dostupná open-source implementace jeví systém dCache (FUHRMANN, 2004). Ten je prověřen nasazením v CERN-u na ukládání velmi velkých objemů dat, jeho nevýhodou je však značná komplikovanost, a tedy vysoké nároky na správce při jeho nasazení.

Závěr

S open-source LTP systémem se počítá v Koncepti rozvoje knihoven. České knihovny tento typ systému potřebují. Mnohé se zabývají bitovou ochranou, ale postrádají efektivní nástroj pro správu archivních dat, který by jim umožnil zvyšovat úroveň

¹⁸ <http://www.lockss.org/community/networks/>

¹⁹ <https://docs.aws.amazon.com/AmazonS3/latest/dev/Welcome.html>

²⁰ <https://cloud.oracle.com/storage>

²¹ <https://du.cesnet.cz/>

²² <https://btrfs.wiki.kernel.org/>

²³ <http://open-zfs.org/>

²⁴ <http://ceph.com/>

²⁵ <https://www.openstack.org/>

jejich logické ochrany. Bitová ochrana je pouze předstupněm k logické ochraně, která má zaručit, že informační obsah uložených dat bude v budoucnu skutečně použitelný (tj. vyhledatelný, pochopitelný a autentický). Pouhé uložení a zálohy dat nezaručí budoucí použitelnost informačního obsahu.

K vývoji systému ArCLib je třeba přistupovat s maximální odpovědností. Pokud má mít takový projekt smysl, musí respektovat specifické potřeby různých typů knihoven. Knihovny budou ArCLib používat v odlišných podmínkách, s různými typy a objemy dat. Systém proto bude muset být poměrně flexibilní. Vývoj systému pro dlouhodobou archivaci je závazek do budoucna. Podobně jako systém Kramerius, i ArCLib bude třeba trvale vyvíjet a udržovat po skončení projektu. Kramerius je dnes po 15 letech existence ve verzi 5 a několikrát prošel zásadními infrastrukturními a funkčními změnami. ArCLib bude systém v pozadí, bude sloužit především několika málo knihovníkům a správcům digitálních dat. Nebude poskytovat klíčové služby přímo čtenářům a uživatelům knihoven jako knihovní katalogy, discovery nebo digitální knihovny. Ovšem pro trvalé udržení služeb knihoven bude neméně důležitý: bez nástrojů pro efektivní správu a ochranu digitálních dat nemusí mít knihovny v budoucnu co čtenářům nabízet.

Použitá literatura

HUTAŘ, Jan a Marek MELICHAR. The long decade of digital preservation in heritage institutions in the Czech Republic: 2002 – 2014. *International Journal of Digital Curation* [online]. 2015, **10**(1), 173-183 [cit. 2016-05-20]. DOI: 10.2218/ijdc.v10i1.324.

KLINDT Marco a Kilian AMRHEIN. *One Core Preservation System for All your Data. No Exceptions!* In: *iPRES 2015*, 2. – 6. 11. 2015, Chapel Hill, USA. Dostupné z: <<https://opus4.kobv.de/opus4-zib/files/5663/iprespaper-finaledit.pdf>>.

MELICHAR, Marek; HUTAŘ, Jan. České paměťové instituce a digitální data – historický exkurz, současný stav a předpokládaný vývoj I. Duha [online]. 2013, roč. 27, č. 4 [cit. 2016-05-25]. Dostupný z: <<http://duha.mzk.cz/clanky/ceske-pametove-institute-digitalni-data-historicky-exkurz-soucasny-stav-predpokladany-vyvoj>>. ISSN 1804-4255.

MELICHAR, Marek; HUTAŘ, Jan. České paměťové instituce a digitální data – historický exkurz, současný stav a předpokládaný vývoj III.. Duha [online]. 2014, roč. 28, č. 2 [cit. 2016-05-25]. Dostupný z: <<http://duha.mzk.cz/clanky/ceske-pametove-institute-digitalni-data-historicky-exkurz-soucasny-stav-predpokladany-vyvoj-1>>. ISSN 1804-4255.

PEJŠOVÁ, Petra. Metodika zpracování, dlouhodobého uchování a zpřístupnění šedé literatury v ČR na příkladu Národního úložiště šedé literatury [online]. Praha: Národní technická knihovna, 2011 [cit. 2016-05-25]. Dostupné z: <http://invenio.nusl.cz/record/111907/files/idr-359_1.pdf>.

RAMALHO, J. C., et al. *RODA and CRiB a service-oriented digital repository*. Communication at International Conference on Preservation of Digital Objects (iPres), London, UK, 2008.

Shepherding the bits. THE ZUSE INSTITUTE BERLIN. *ZIB* [online]. c2014 [cit. 2016-26-05]. Dostupné z: <<http://www.zib.de/features/shepherding-bits>>.

FUHRMANN, Patrick. dCache, the Overview. White paper, 2004. [cit. 2016-05-27]. Dostupné z: <<https://www.dcache.org/manuals/dcache-whitepaper-light.pdf>>.

Mgr. Marek Melichar

marek.melichar@ruk.cuni.cz

(Ústav výpočtové techniky Univerzity Karlovy v Praze)

Mgr. Jan Hutař

jan.hutar@dia.govt.nz

(Archives New Zealand)

RNDr. Michal Růžička

mruzicka@mail.muni.cz

(Ústav výpočtové techniky Masarykovy univerzity v Brně)

Bc. Zdeněk Hruška

Zdenek.Hruska@mzk.cz

(Moravská zemská knihovna v Brně)

Mgr. Zdeněk Vašek

zdenek.vasek@nkp.cz

(Národní knihovna České republiky)

RNDr. Miroslav Bartošek, CSc.

bartosek@ics.muni.cz

(Ústav výpočtové techniky Masarykovy univerzity v Brně)

Ing. Martin Lhoták

lhotak@knav.cz ■

(Knihovna akademie věd České republiky)