**Příloha č. 1 – Technická specifikace předmětu zakázky na rozvoj systému ARCLib**

**POPIS PROJEKTU**

Archivační systém ARCLib byl jako open source software vyvinut v letech 2016-2020 s finanční podporou Ministerstva kultury ČR v rámci národního programu financování aplikovaného výzkumu NAKI II. Pro realizaci projektu se spojily Masarykova univerzita, Národní knihovna ČR a Moravská zemská knihovna v Brně do konsorcia, které zastřešovala a vedla Knihovna AV ČR, v. v. i. Tyto instituce nyní realizují většinu aktivit spojených s dlouhodobým uchováváním digitálních dokumentů v knihovnách v ČR.

Projekt ARCLib reagoval na potřeby paměťových institucí, zejména knihoven, zajistit dlouhodobé uchování digitálních dokumentů. Součástí projektu byla příprava metodických materiálů a technických řešení, které jsou nyní volně k dispozici. Kromě softwarového řešení ARCLib, které je včetně dokumentace dostupné na https://github.com/LIBCAS/ARCLib – jde o certifikované metodiky s doporučeným postupem pro bitovou a logickou ochranu digitálních dokumentů v systému ARCLib (viz http://hdl.handle.net/11104/0294749 a http://hdl.handle.net/11104/0282107).

Vývoj volně dostupného řešení pro dlouhodobou archivaci, které je k dispozici knihovnám v České republice, významně přispívá naplňování Koncepce rozvoje knihoven České republiky na roky 2021-2027, kde je v rámci klíčové oblasti III. Digitální fondy a data definován cíl: "Zajistit provoz a rozvoj infrastruktury LTP úložišť v NK a největších knihovnách." Vývoj alternativního volně dostupného LTP řešení byl explicitně zmíněn také v předchozí koncepci platné do roku 2020.

V současné době je systém ARCLib nasazen v ostrém provozu v Knihovně Akademie věd ČR. Mezi zájemce o provozování patří např. Ústřední knihovna Univerzity Karlovy. Zprovoznění je plánováno i v Moravské zemské knihovně a na Masarykově univerzitě. Další zájemci mají možnost požádat o dotaci na implementaci v rámci dotačního programu VISK7.

Systém ARCLib splňuje základní předpoklady pro využití jako LTP systém. Aby splnil širší nároky uživatelů v kontextu českých paměťových institucí, především knihoven, je však třeba vyvinout další funkce, které umožní realizovat kompletní workflow pro správu digitálních dat a jejich dlouhodobé uchovávání dle konceptu OAIS. Dotace z programu VISK poskytnuté na rozvoj ARCLibu v letech 2022 a 2023 umožnily rozšíření možností exportu a importu, rozšíření vyhledávání, mazání a aktualizace archivních balíčků a update knihoven, které jsou v systému využívány, což je velmi důležité i z hlediska zachování bezpečnosti celého řešení. Na základě analýz projektového týmu, složeného ze zástupců institucí působících v původním NAKI projektu (viz výše), provedených v roce 2023, došlo k návrhu rozšíření ARCLibu o další funkce, které přispějí ke kvalitě systému, širším možnostem správy archivovaných dat a jednodušší implementaci pomocí Dockeru.

Zdrojový kód a dokumentace je k dispozici na <https://github.com/LIBCAS/ARCLib>

Rozvoj systému ARCLib v roce 2024 – předmět veřejné zakázky:

**Export vybraných záznamů z hromádky**

Uživatel má možnost odkládat výsledky vyhledávání na hromádku a následně ji jako celek exportovat, v současnosti nelze exportovat pouze podmnožinu hromádky.

**Export více druhů identifikátorů**

Export identifikátorů v současnosti umožňuje pouze export XML ID (ARCLIB\_000..XX).

**Ukládání uživatelských preferencí tabulek**

V současnosti je nastavení filtrů a řazení tabulky ztraceno při opuštění tabulky. Kromě uložení preferencí bude dostupná i možnost obnovení výchozího nastavení.

**Rozšířená konfigurace tabulek**

V současnosti je dostupné jednoúrovňové řazení a filtrování. Mohli by přibýt možnosti definice víceúrovňového řazení (např. nejprve dle názvu, následně dle datumu), možnost vybrat které sloupce jsou zobrazeny, možnost seřadit sloupce dle potřeby (např. název první). I tato nastavení by se ukládala a bylo by možné je resetovat.

**Export tabulek do CSV/XLSX**

Tabulky (v podobě dle zvolených preferencí konkrétního uživatele, viz výše) by bylo možné exportovat do CSV/XLSX souboru. V případě tabulek se stránkováním by uživatel volil zda má být ve výstupu aktuálně zobrazená stránka, nebo všechny stránky.

**Více forem autentizace**

Současný ARCLib vyžaduje k autentizaci LDAP server. Jako alternativu by bylo vhodné poskytnou i možnost autentizace přímo vůči interní DB. V rámci implementace budou na vhodných místech použita rozhraní, tak aby zbytek systému byl odstíněn od konkrétní implementace a bylo možné v budoucnu případně zavést i další způsoby autentizace.

Jedná se pouze o prvotní autentizaci k získání JWT tokenu, autorizace a řízení expirace sezení by zůstala pro všechny druhy autentizace v současném stavu.

**Dockerizace**

Současná instalace ARCLibu je z důvodu velkého množství služeb poměrně náročná. Dockerizace by umožnila potenciálním zájemcům rychlejší seznámení se systémem, pro menší archivy by mohly být některé kontejnery i součástí produkční infrastruktury. V rámci dockerizace by byly implementovány následující kontejnery:

• PostgreSQL

• Solr

• Arclib (společně s aplikací Droid)

• Archival Storage

• ClamAV

• LDAP (pokud nebudou implementovány jiné formy autentizace)

**Automatický reingest**

Při vydání nové verze ARCLibu s zásadními změnami např. ve formátu ARCLib XML, může být pro podporu všech funkcionalit potřebné provést kompletní reingest současného obsahu. Aktivita reingestu v současnosti není automatizovaná, subjekt provádějící reindex musí nyní získat kompletní obsah archivu (ať už z Archival Storage nebo z jiných, vlastních úložišť) a vložit ho znovu na vstup (rozřadit do správných složek ve workspace) ARCLib ingestu. Automatizovaný reingest by toto provedl automaticky: balíčky, které jsou v době spuštění trvale archivovány (stav PERSISTED) by postupně extrahoval z Archival Storage do speciálních složek pro reingest a spouštěl jejich nový ingest.