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

ProArc je open-source systém umožňující tvorbu popisných, technických a administrativních metadat pro digitalizované i born-digital dokumenty. V současné době umožňuje popis periodik, monografií, zvukových dokumentů (gramofonových desek i fonografických válečků), elektronických periodik, elektronických monografií a starých tisků včetně konvolutů.

ProArc byl vyvinut v rámci projektu Česká digitální knihovna, financovaného z programu MK ČR NAKI v letech 2012-2015. K financování dalšího průběžného vývoje ProArcu byl využit také následný NAKI projekt ARCLib a částečně také program VISK. Vývoj systému od samého začátku koordinuje Knihovna Akademie věd ČR a její Digitalizační centrum. Jednou z prvních knihoven, která se k vývoji připojila, byla Studijní a vědecká knihovna v Hradci Králové. Součástí vývojového týmu jsou v současnosti také Moravská zemská knihovna v Brně (MZK) a Městská knihovna v Praze. Z průzkumu, který provedl na jaře 2023 vývojový tým mezi uživateli ProArcu, vyplynulo, že systém k digitalizaci pravidelně využívá 14 institucí a v roce 2022 tak pomocí něho bylo zpracováno přibližně 4600 monografií, 1250 ročníků periodik a 750 zvukových dokumentů.

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

Vývojové práce v roce 2024 – předmět veřejné zakázky:

**Implementace aktualizovaných Definic metadatových formátů NDK**

Digitální dokumenty zpracované v systému ProArc jsou snadno publikovatelné v systému Kramerius, který pro provoz své digitální knihovny využívá více než 40 knihoven v ČR. Vytvořené digitální dokumenty plně respektují Definice metadatových formátů (DMF) pro jednotlivé typy dokumentů vydávané Národní knihovnou ČR. ProArc tedy umožňuje tvorbu digitálních dokumentů v souladu s podmínkami programu VISK 7 a zároveň produkuje data vhodná pro dlouhodobé uchování kompatibilní s LTP systémem Národní knihovny ČR nebo open-source systémem ARCLib. V květnu 2023 došlo ke schválení nových verzí DMF pro Periodika a Monografie. Významnou změnou je zejména podpora mutačních vydání čísel periodik, která umožňuje popsat digitalizované dokumenty s regionálními odlišnostmi. V roce 2023 se věnovala Městská knihovna v Praze ve spolupráci s Oddělením standardů digitálních sbírek Národní Knihovny přípravě DMF pro zpracování výstřižků a KNAV se spolu s MZK věnovala návrhu aktualizace metodického pokynu pro digitalizaci starých tisků v rámci nástrojů pro digitalizaci novodobých dokumentů. U obou těchto dokumentů se předpokládá dokončení v průběhu roku 2024.

Aby byla dále umožněna produkce digitálních dokumentů v souladu s platnými standardy, je třeba implementovat všechny tyto změny do systému ProArc. Součástí implementace bude i možnost vytvořit rozšířený archivní balíček pro zvukové dokumenty.

**Born-digital dokumenty**

Téměř třetina institucí, které ProArc používají, v něm zpracovává born-digital dokumenty. Lze předpokládat, že jejich počet se bude dále zvyšovat. Pro dlouhodobé uchování těchto dokumentů je zásadní jejich převod do adekvátního archivního formátu. K zefektivnění práce s nimi bude do ProArcu implementována funkce pro automatický převod PDF dokumentů do archivního formátu PDF/A. Současně bude umožněno převést born-digital dokumenty vytvořené před vznikem adekvátního standardu (před rokem 2021) do podoby respektující v současnosti platné znění standardu. Rozšířena bude také možnost exportu metadat vědeckých článků do mezinárodní databáze Crossref, která nově umožní zapisovat k jednotlivým článkům i seznam použité literatury, což vylepší možnost jejich nalezení, sdílení a využívání. Tyto informace budou zároveň zapsány i do popisných metadat ve schématu MODS.

**Rozvoj klientského rozhraní**

Systém ProArc se skládá z jádra disponujícího REST API a klientské aplikace, kterou pro svou práci používají jednotliví zpracovatelé. V průběhu roku 2023 byly identifikovány a popsány návrhy na úpravu klientského prostředí, které umožní rychlejší a efektivnější zpracování digitálních dokumentů a kontrolu zpracování a procesů v systému. Tyto návrhy spadají do dvou kategorií. Adekvátně k těmto změnám bude upravena i komponenta RDFlow sloužící k monitorování a řízení některých pracovních procesů.

* Navigace v rozhraníHlavním návrhem v této oblasti je úprava podoby obrazovky „Úložiště“ tak, aby byly dokumenty zobrazeny ve stromové hierarchické struktuře. Ta umožní snazší a rychlejší navigaci po jednotlivých úrovních zpracovávaných dokumentů bez nutnosti jejich postupného otevírání. Současně budou přímo ve struktuře stromu zobrazována některá metadata dokumentů, včetně počtu stran v rámci jednotlivých titulů pro rychlý přehled a kontrolu.
* Kontrolní mechanismy
Lepší kontrolu zpracovávaných dokumentů zajistí úpravy zahrnující další vybrané funkce. Pro včasné odhalení chyb bude sloužit automatické rozpoznávání duplicit vybraných typů stran (přední desky, titulní strana, zadní desky,...), automatická kontrola některých hodnot ve vybraných metadatových elementech a celková validace před spuštěním exportu dat. Z hlediska kontroly procesů probíhajících v aplikaci bude nově možné pozastavit a případně zrušit již probíhající exportní procesy. Na úspěšné dokončení procesu importu dat z ProArcu do digitální knihovny Kramerius bude navázáno automatické doplnění odkazu na digitalizovaný dokument do knihovního katalogu. Ošetřeno bude i vypršení platného přihlášení dlouho neaktivního uživatele. Součástí kontrolních funkcí bude také volitelná barevnost aplikace na úrovni systémové konfigurace, která zajistí v institucích s více instalacemi ProArcu jasné rozlišení mezi testovacími a produkčními instancemi. Aplikace také umožní odstranění uživatelů ze systému se zachováním jimi vytvořených objektů.

Výše uvedené rozvojové návrhy je nutné realizovat, aby bylo nadále zajištěno dodržování Definic metadatových formátů vydávaných Národní knihovnou ČR a umožněna interoperabilita dat napříč archivačními a publikačními systémy.