**Příloha č. 2 – Technická specifikace předmětu zakázky na Rozvoj produkčního systému ProArc v roce 2023**

**Rozbor stávající situace:**

Open source systém ProArc slouží k tvorbě popisných, technických a administrativních metadat k digitalizovaným i born-digital dokumentům v souladu se standardy NDK. Postupně jsou rozšiřovány typy objektů, které lze popisovat – aktuálně je možný popis periodik, monografií, starých tisků, článků, elektronických publikací a gramofonových desek a fonografických válečků. ProArc zároveň umožňuje vytvoření archivačního balíčku, který může být následně archivován v LTP systému ARCLib.

ProArc byl vyvinut zejména díky Digitalizačnímu centru Knihovny AV ČR v Jenštejně v rámci projektu Česká digitální knihovna financovaného z programu MK ČR NAKI. Vývoj zajištěný z tohoto projektu byl dokončen v roce 2015. K financování dalšího průběžného vývoje ProArcu byl využit také následný NAKI projekt ARCLib a částečně také program VISK. Jednou z prvních knihoven, která začala na vývoji systému ProArc spolupracovat, byla Studijní a vědecká knihovna v Hradci Králové. Postupně se do vývojového týmu připojily i další knihovny – Moravská zemská knihovna v Brně, Městská knihovna v Praze či Národní muzeum. Další instituce nejsou přímo zapojeny do vývojového týmu, ale systém ProArc používají, jedná se například o Ústřední knihovnu Univerzity Karlovy nebo Knihovnu A. Švehly (ÚZEI) a objevují se i další zájemci o práci s tímto systémem.

Digitální dokumenty zpracované v systému ProArc jsou primárně publikovány v systému Kramerius a plně respektují definice metadatových formátů pro jednotlivé typy dokumentů vydávané Národní knihovnou ČR. Na konci roku 2022 došlo k významné aktualizaci těchto standardů, a aby byla umožněna produkce digitalizovaných dat v souladu s jejich novými verzemi, je třeba zajistit implementaci všech změn do systému ProArc. Spolu s těmito změnami budou provedeny i přípravné práce pro zavedení nově vznikajícího standardu pro výstřižky a lepší editace technických metadat pro born-digital dokumenty. Za zásadní úkol úzce související s definicemi metadatových formátů považuje vývojový tým také vytvoření vhodných podmínek pro úpravy dříve digitalizovaných dokumentů, které byly zpracovány v již nepoužívaných produkčních systémech a v době před vyhlášením standardů NDK. Tyto dokumenty jsou často uloženy pouze v jednotlivých instancích systému Kramerius a možnost efektivně převést tyto dokumenty z Krameria do ProArcu by umožnilo jejich obohacení o často úplně chybějící textovou vrstvu OCR/ALTO a opravu metadat, díky které by byla zajištěna kompatibilita těchto starých dat s aktuálně platnými standardy NDK. Navíc se tak otevře i cesta k vytvoření archivačních balíčků splňujících standardy pro digitalizaci a dlouhodobé uchování, které budou kompatibilní s open source LTP systémem ARCLib i s LTP systémem Národní knihovny v lince NDK.

Během roku 2022 se vývojový tým ProArcu věnoval mimo jiné užšímu, resp. více automatizovanému propojení systémů Kramerius a ProArc a v rámci I. fáze tohoto propojení byla realizována možnost automatické aktualizace metadat dokumentů v Krameriovi zápisem přímo z ProArcu přes API a částečná možnost zpětného reimportu dat z Krameria do ProArcu. V roce 2023 naváže rozšíření komunikace obou systémů II. fází, která má zajistit další zefektivnění a automatizaci činností spojených s publikováním nově vytvořených digitálních dokumentů. Díky užšímu propojení obou systémů bude možno na základě navazujících procesů v Krameriovi automaticky ovlivňovat kroky workflow (RDFlow) integrovaného v ProArcu. Důležitým krokem s vazbou mj. na správu dokumentů v režimu DNNT bude umožnění volby parametrů zveřejnění dokumentů již v ProArcu včetně zápisu údajů o licenci do metadat exportního balíčku a jejich odeslání do Krameria.

Adekvátně k plánovaným změnám v systému ProArc bude probíhat také rozvoj klientské aplikace, kterou pro svou práci používají jednotliví zpracovatelé dokumentů. Součástí dodaných prací je údržba a správa projektu na veřejném portálu na https://github.com/proarc (zveřejnění aktuálního zdrojového kódu, komunikace k issues, projektová dokumentace,…)

Program VISK je v tuto chvíli jediným možným zdrojem účelového financovaní pro rok 2023, přičemž řada z výše uvedených vývojových kroků musí být realizována v souvislosti s dodržováním standardů, další pak urychlí a zefektivní práci při produkci a úpravách digitálních dokumentů určených pro zpřístupnění v systému Kramerius.

**Požadavky na vývoj a služby, které jsou předmětem zakázky:**

**Aktualizace a úpravy související s novými standardy NDK**

* Zavedení nových standardů do ProArcu
* Definice metadatových formátů pro digitalizaci periodik 1.9
* Definice metadatových formátů pro digitalizaci monografických dokumentů 2.0
* Přípravné práce pro implementaci/příp. implementace standardu pro výstřižky
* Možnost editace technických metadat pro born-digitaldokumenty
* Rozšíření komunikace s resolverem (opravy urn:nbn)
* Vytvoření podmínek pro úpravy starších digitalizovaných dokumentů, aby odpovídaly platným standardům
* zpětné importy dat z Krameria
* importy NDK SIP balíčků
* tvorba ALTO/OCR ke starým digitalizacím (tj. např. k FOXML z Krameria naimportovaného do ProArcu)

**II. fáze užšího propojení mezi systémy ProArc a Kramerius**

* Import NDK SIP do nakonfigurované instance Krameria
* Příprava ProArcu na práci s licencemi Krameria
* zápis licence do metadat
* možnost v rámci importu zvolit zvolit parametry zveřejnění dokumentu v digitální knihovně
* RDflow - spolu s užším propojováním mezi systémy ProArc a Kramerius bude dále rozvíjena komponenta RDFlow

**Rozvoj klienta a správa repozitářů ProArcu**

* Zobrazení indexů a zkrácených nebo celých popisků stran v návaznosti na nové možnosti systému postaveného na Akubře
* Vytváření NDK čísel periodika přes funkci "více"
  + s předáním signatury podřízeným úrovním
  + s možností zadat speciální datování/periodicitu
* Připojení původních nezpracovaných skenů k archivnímu balíčku
* Označování left/right/single page + reprezentativní strany u STT stran
* Správa repozitářů ProArcu na GitHubu <https://github.com/proarc/proarc> a doplnění dokumentace ke všem částem nově vyvinutého řešení