



**Popis poloprovozu vytváření
a zpřístupnění digitálních archivních
balíčků se zaměřením na umělecká
díla s audiovizuálním obsahem
nekinematografické povahy**

1. Digitalizační fáze	4
1.1. Popis digitalizačního řetězce	4
1.2. Hardwarové řešení digitalizační fáze	5
1.3. Softwarové řešení digitalizační fáze	7
1.4. Postup při digitalizaci páskových nosičů	7
1.5. Výběr vhodného zdroje dle významných vlastností	9
2. Archivační fáze	11
2.1. Hardwarové řešení archivační fáze	12
2.2. Softwarové řešení archivační fáze	12
2.2.1. Transfer	13
2.2.2. Ingest	14
2.2.3. Archival Storage	14
2.3. Technická specifikace archivačních formátů	15
3. Prezentační fáze	16
3.1. Popis přístupového portálu	16
3.2. Hardwarové řešení přístupového portálu	16
3.3. Softwarové řešení přístupového portálu	17
4. Ekonomické parametry poloprovozu	18
5. Přílohy	19
5.1. Vzor konzervátorské zprávy	19
5.2. Soupis technického vybavení	23

Autor: Jonáš Svatoš, Národní filmový archiv

Verze dokumentu: 2022-03-03

Projekt: NAKI II DG20P020VV025 “Audiovizuální dílo mimo kontext kinematografie: dokumentace, archivace a zpřístupnění”

Poloprovoz, který je výsledkem projektu *Audiovizuální dílo mimo kontext kinematografie: dokumentace, archivace a zpřístupnění DG20P020VV025* financovaný z programu NAKI II Ministerstva kultury ČR, je komplexním řetězcem pro zpracování audiovizuálních objektů určených k archivaci. Skládá se z akviziční přípravy, vytvoření vstupního balíčku, jeho zpracování dle modelu OAIS a prezentace v odborné databázi. Vlastníkem výsledku je Národní filmový archiv.

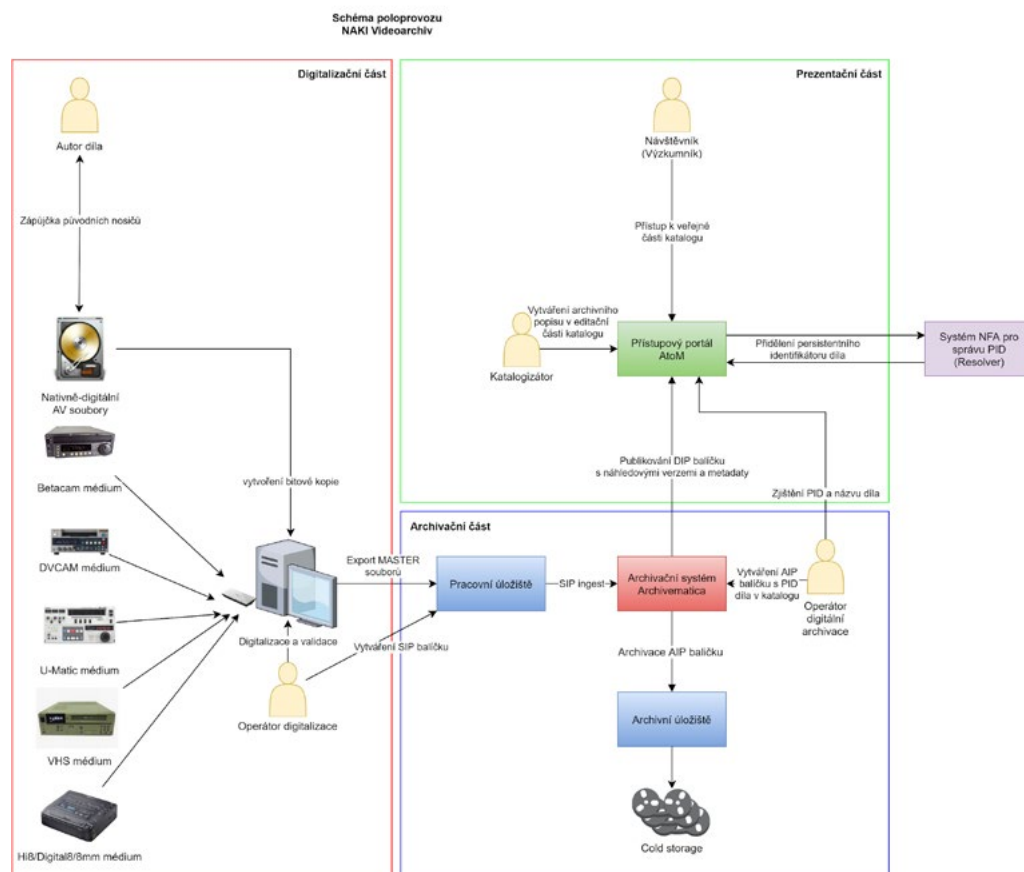
Řetězec sestává z hardwarových komponent, na kterých je provozována softwarová akviziční a archivační platforma. Akviziční část zajišťuje digitalizaci širokého spektra elektronických obrazových formátů (VHS, Hi-8, Betacam SP, Digital Betacam, miniDV, DVCAM, U-matic, CD-ROM, DVD). Realizuje i příjem nativně digitálních médií. Část sestavení a zpracování archivního balíčku ukládá data jednotlivých objektů dle modelu OAIS. Společně s daty je ukládán jejich metadatový popis ve standardu PREMIS. Část prezentace archivních objektů zpřístupňuje uložené informace pro kurátorskou praxi a další užití s možným předáním dat do registru digitalizace.

Poloprovoz se skládá ze tří fází:

- digitalizace nosičů a uložení dat
- archivace – vytváření archivních balíčků a prezervačních metadat
- prezentace archivních objektů v přístupovém portálu, který je rozhraním k databázi

Následující schéma ukazuje provázanost všech tří částí řetězce. Červená část znázorňuje fázi akvizice, modrá fázi archivace, zelená prezentační fázi. Následující kapitoly podrobně rozebírají jednotlivé fáze především po technologické stránce, včetně vzájemné návaznosti.

Funkční schéma poloprovozu pro digitalizaci a archivaci multimediálních nosičů



1. DIGITALIZAČNÍ FÁZE

Předpokladem pro započetí procesu akvizice je dostupnost konkrétních audiovizuálních objektů (ať již fyzických, či nativně digitálních), které lze použít jako vstup do poloprovozu. Dalším předpokladem je, že vhodné kandidátní objekty odhalí kurátoři, na základě historického výzkumu, během identifikační projekce a v rozhovoru s původci děl.

Akviziční část poloprovozu se tedy především zabývá pořízením digitálních kopií audiovizuálních záznamů z původních fyzických nosičů či z datových souborů a jejich technickým popisem. Tento proces zajišťuje digitalizační řetězec, který se skládá z fyzických zařízení schopných věrně reprodukovat původní záznam na nosičích obsažených, dále pak sady softwarových nástrojů umožňujících automatizovaný sběr technických metadat z digitalizátů. Ty dále umožňují sledovat kvalitativní parametry záznamu, na jejichž základě lze rozhodnout o případných úpravách parametrů digitalizace.

1.1. Popis digitalizačního řetězce

Digitalizační řetězec byl navržen tak, aby bylo umožněno v průběhu práce měnit rozložení v závislosti na potřebě digitalizace nových formátů. Základ tvoří profesionální videopřehrávače kazetových formátů, které byly vybrány v souladu s doporučeními technické komise Mezinárodní asociace zvukových a audiovizuálních archivů (IASA TC), konkrétně normy IASA TC-06,¹ která byla výchozím zdrojem při vyměřování celého poloprovozu. Neméně důležitou součástí tvoří dostatečně výkonnostně dimenzované digitalizační stanice založené na všeobecně dostupných hardwarových komponentách, které je možno do budoucna obnovovat dle potřeby. Digitalizační rozhraní bylo vybráno taktéž dle doporučení IASA TC-06, zejména vzhledem ke kvalitě a rozlišení A/D převodníků, které jsou stěžejním bodem celého řetězce. Pro funkci digitalizačního software bylo zvoleno otevřené řešení, které je volně dostupné včetně zdrojového kódu.

Záměrem digitalizační části je získat co možná nejvěrnější digitální otisk děl obsažených na původním fyzickém nosiči. Předpokladem však je, že takto vzniklá faksimile nikdy nebude přesnou kopií originálu vzhledem k tomu, že dochází k postupnému zlepšování technických možností digitalizace. Lze tedy prohlásit, že zde navržený postup využívá ty nejlepší běžně dostupné technologické postupy vzhledem k současnému stavu poznání na tomto poli. Zároveň je však třeba mít na paměti, že trvanlivost záznamu na audiovizuálních nosičích je omezená, a proto je třeba jednat bezodkladně, i když je pravděpodobné, že v budoucnosti budou k dispozici ještě dokonalejší technické nástroje. Existuje však reálné riziko, že v tomto časovém horizontu předmětné audiovizuální nosiče již nebudou obsahovat dostatek informací k rekonstrukci původního díla kvůli k jejich postupné fyzické degradaci.

¹ IASA-TC 06 - Guidelines for the Preservation of Video Recordings, cit. , <https://www.iasa-web.org/tc06/guidelines-preservation-video-recordings>.

1.2. Hardwarové řešení digitalizační fáze

Parametry digitalizačního rozhraní byly zvoleny na základě doporučení IASA TC-06, a byly stanoveny následovně:

Pro video:

bitová hloubka	Chroma subsampling	Barevné prostory	Vstupní rozlišení	Vstupní snímková frekvence
min. 10-bit/kanál	4:4:4, 4:2:2, 4:2:0	Rec.601, Rec.709	PAL, NTSC, 720p, 1080i, 1080p	24p, 25p, 29.97p, 30p, 50p, 59.94p, 60p, 50i, 59.94i, 60i, 23.98PsF, 24PsF, 25PsF, 29.97PsF, 30PsF

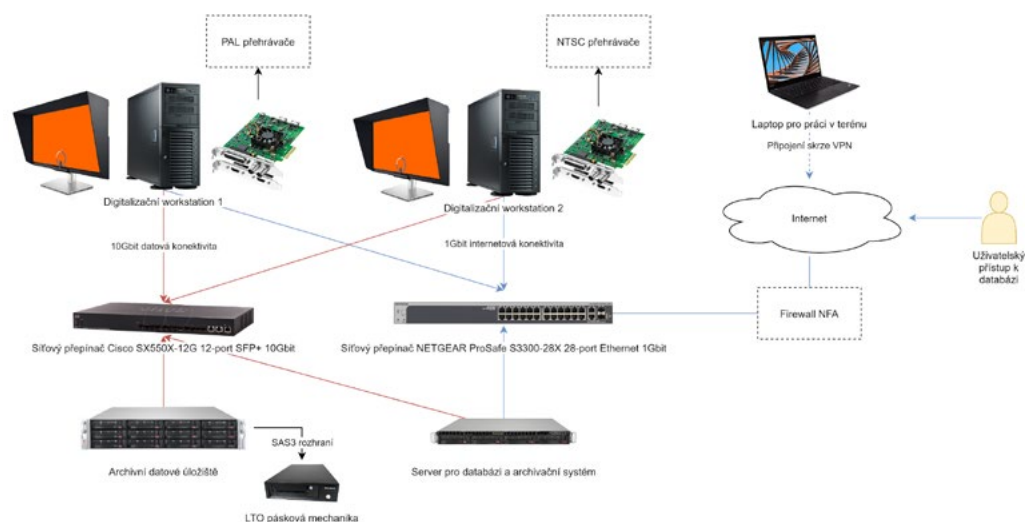
Pro audio:

bitová hloubka	Samplovací frekvence
min. 24-bit/kanál	min. 48 kHz

Seznam technického vybavení digitalizačního pracoviště je součástí přílohy č. 3. Pro přehlednost neuvádíme majetek, jako jsou sluchátka, zásuvky, kabely, zdroje nepřerušovaného napájení apod.

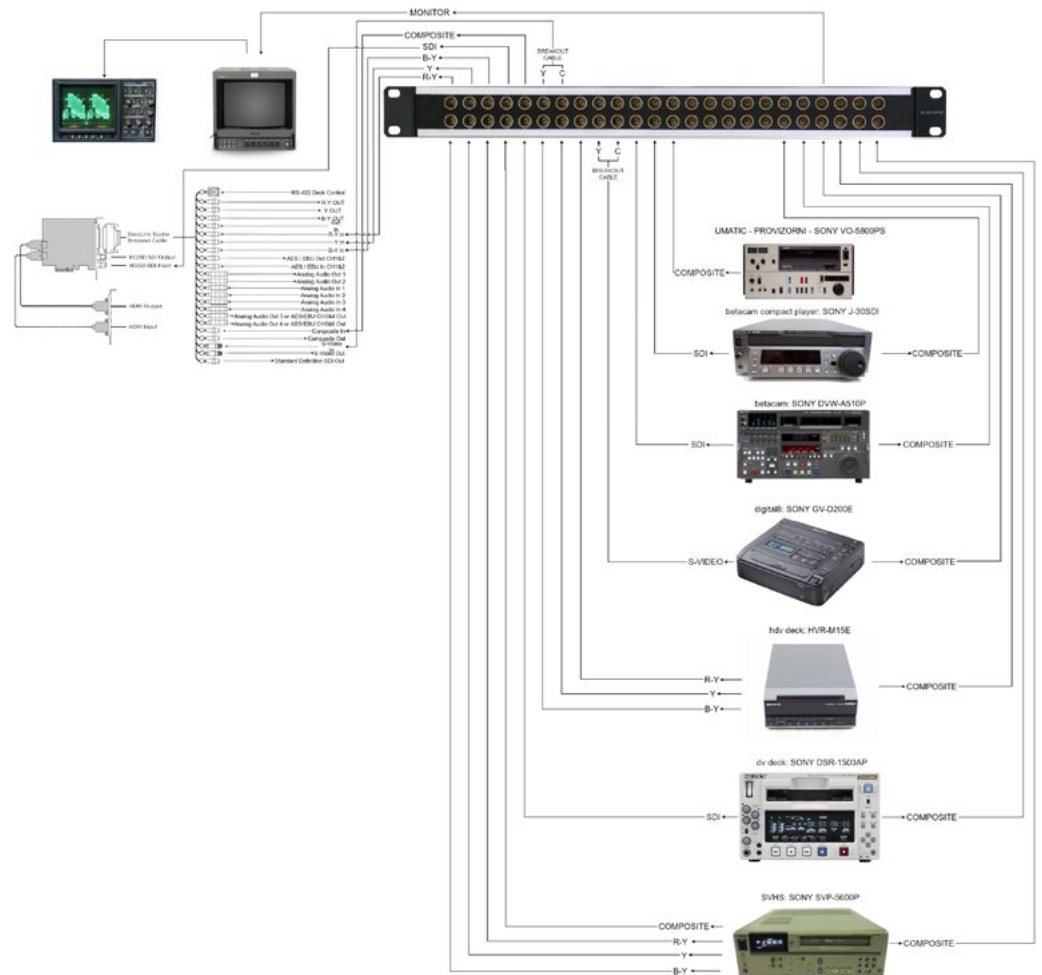
Následující schémata znázorňují vzájemné propojení úložné infrastruktury a digitalizační techniky:

Schéma IT infrastruktury Videoarchivu



IT infrastruktura je fyzicky umístěna na pracovišti Videoarchivu v archivním areálu Národního filmového archivu v obci Hradištko, Brunšov. Diagram znázorňuje, jak jsou pracovní stanice a servery síťově propojeny. Infrastruktura má dva oddělené síťové okruhy – jeden pro přístup do internetu s rychlostí 1 Gbit/s a druhý pro datovou konektivitu s rychlostí 10 Gbit/s, který je od internetu fyzicky oddělen. Obě digitalizační stanice jsou připojeny do obou sítí, a to zejména kvůli nutnosti přistupovat jak na datové úložiště Videoarchivu, tak na ostatní interní systémy NFA. Ty jsou provozovány v datovém centru v Praze, kde taktéž sídlí centrální firewall, který umožňuje přístup z a do internetu. Server pro aplikace, kde běží veřejná databáze i archivní systém, je připojen taktéž do obou sítí kvůli přístupu uživatelů z internetu a zároveň nutnosti zpracovávat data na úložišti. Archivní datové úložiště je připojeno pouze do datové sítě z důvodů bezpečnosti. K datovému úložišti je připojena LTO pásková mechanika, pomocí které se výsledné archivní balíčky ukládají na dvě oddělené sady LTO7 datových médií (o tom více v kapitole 2. 1.).

Schéma propojení digitalizační techniky



Propojení jednotlivých přehrávačů a ostatní digitalizační techniky znázorňuje diagram výše. Na něm je patrné, že propojení je modulární díky propojovacímu panelu. Ten operátorovi umožňuje dynamicky měnit přehrávače podle toho, jaký je k digitalizační kartě připojen pomocí krátkých propojovacích kabelů (tzv. patch kabelů). Ty se zapojují mezi dvě zdičky tak, aby vytvořily spojený okruh. Zapojení digitalizační karty i přehrávačů do panelu je tedy fixní, operátor mění propojení pouze pomocí patch kabelů.

Fotografie níže ukazuje jedno ze dvou digitalizačních stanovišť Videoarchivu. Obě stanoviště jsou identická, liší se pouze televizní normou jednotlivých přehrávačů (pracoviště 1 pro PAL, pracoviště 2 pro NTSC).



1.3. Softwarové řešení digitalizační fáze

Softwarová platforma akvizice je založena na operačním systému Rocky Linux 8, binárně kompatibilním sestavení distribuce Red Hat Enterprise Linux (RHEL) 8, dostupné komerčně s dlouhodobou podporou. Jedním z důvodů použití dlouhodobě podporované stabilní verze systému s neměnným aplikačním a systémovým rozhraním (API/ABI) je digitalizační rozhraní firmy Blackmagic (viz výše), jehož softwarové rozhraní vyžaduje konkrétní verzi jádra systému.

Ačkoliv výrobce Blackmagic k digitalizačnímu rozhraní dodává vlastní software pro digitalizaci (Blackmagic Media Express), z důvodu absence potřebných funkcionalit (především podpora digitalizace do bezztrátových kodeků a vkládání uživatelských metadat) jsme se rozhodli použít otevřené řešení “Vrecord”,² které bylo vyvinuto sdružením AMIA³ (Association of Moving Image Archivists) ve Spojených státech. Toto řešení je multiplatformní (podpora macOS a Linux) a umožňuje profesionální úroveň akvizice a kontroly výstupu digitalizátů, včetně vkládání metadat o průběhu procesu digitalizace. V rámci projektu byl vyvinut předpis pro vytvoření softwarového balíčku pro RHEL 8 a jeho derivátů, včetně Rocky Linux.

Pro postprodukcí digitalizátů je použit software Blackmagic Resolve 17 Studio. Bohužel tento ani prozatím žádný z profesionálních stříhových programů neumožňuje přímé dekódování videokodeku FFv1, proto je třeba před editací vždy provést převod obsahu do nekomprimované podoby (konkrétně kodek Uncompressed 4:4:4 10-bit, který je taktéž označován jako “v410”). Pro zachování kvality je nutno výsledný výstup následně taktéž exportovat do tohoto formátu a posléze pomocí transkodéru FFmpeg ho opět zkomprimovat do bezztrátového FFv1.

Vzhledem k rozšířenosti použití tohoto kodeku v audiovizuálních institucích a jeho nedávné standardizaci skupinou IETF⁴ lze předpokládat, že výrobci profesionálních studiových řešení podporu tohoto kodeku do svých řešení brzy implementují.

1.4. Postup při digitalizaci páskových nosičů

Postupy při digitalizaci páskových nosičů jsou založeny na doporučeních obsažených v IASA TC-06, obsažených především v části B, která se zabývá cílovými formáty.

Postup digitalizace:

1. Médium je nejprve zapsáno do tabulky evidence materiálů, ve které je médiu přiřazen identifikátor, pod kterým se s médiem pracuje do doby ukončení akvizice a vrácení nosiče původnímu vlastníkovi.
2. Pořízení fotografií fyzického média pro zachycení informací přímo na nosiči obsažených.

2 Vrecord, cit. 7. 3. 2022, <https://github.com/amiaopensource/vrecord>.

3 Association of Moving Image Archivists, cit. 7. 3. 2022, <https://amianet.org/>.

4 FFV1 Video Coding Format Versions 0, 1, and 3, cit. 7. 3. 2022, <https://data-tracker.ietf.org/doc/rfc9043/>.

3. Ohledání média, zdali jeho fyzický stav umožňuje digitalizaci. V případě, že ne, operátor ve spolupráci s kurátorem zjišťuje existenci dalších zdrojových nosičů.
4. Pokusné přehrání média v jednom z přehrávačů dle formátu a inspekce obrazové kvality média, která probíhá vizuálně, a pokud jsou na začátku média tzv. SMPTE Color Bars, tak dle nich na připojeném osciloskopu.⁵
5. Pokud to video signál vyžaduje, na základě obrazové inspekce operátor provede úpravu nastavení přehrávače (kontrast, saturace, případně použití Timebase korektoru) s cílem dostat obrazový a zvukový signál do normy (v případě přítomnosti tzv. test-barů objektivně, bez nich subjektivně, výsledná podoba digitalizátu je autorizována, viz níže).
6. Operátor na digitalizační stanici v software Vrecord nastaví parametry záznamu.
7. Operátor v software Vrecord započne nahrávání a spustí přehrávání média.
8. Během digitalizace operátor sleduje obraz na připojeném CRT TV monitoru a výsledný digitalizát na plochém monitoru. Obdobně sleduje připojené analogové i digitální osciloskopy, aby případně zachytil jakékoliv výpadky v signálu způsobené např. fyzickými nečistotami na nosiči.
9. Digitalizát je uložen na lokálním pracovním úložišti do adresáře s názvem díla a jeho identifikátorem.
10. Digitalizát je poskytnut autorovi či jinému původci pro autorizaci. Pokud je digitalizát schválen, stává se z něho tzv. Autorizovaná reprezentace původního díla a zůstává na lokálním pracovním úložišti, kde z něj operátor (spolu s ostatními vybranými verzemi díla) následně vytváří SIP balíček pro vstup do procesu archivace (o tom více ve fázi 2).

V případě, že je na nosiči jedno dílo vícekrát (např. pro potřeby výstavního provozu), provádí se digitalizace v plné délce a operátor následně provede identifikaci nejvhodnější varianty vzhledem k případné přítomnosti různých obrazových vad, kazů na pásu samotném apod. V případě, že nelze spolehlivě určit, jaká varianta je nejvhodnější, je přípustné provést sestavení optimální varianty z několika po sobě jdoucích kopií v digitální střížně.

⁵ Postup kalibrace dostupný v dokumentaci Vrecord, cit. 7. 3. 2022, https://github.com/amiaopensource/vrecord/blob/main/Resources/Documentation/analog_digitization.md.

1.5. Výběr vhodného zdroje dle signifikantních vlastností

Posouzením technických charakteristik musí projít každý z fyzických nosičů a digitálních souborů zamýšlených pro akvizici Videoarchivu. Dokument se vztahuje na všechny materiály elektronické povahy i digitální akvizice. Cílem posouzení je zejména:

- **Potvrzení identity materiálu.** – Jde o očekávaný typ obsahu vzhledem k vybraným předpokladům? Nakolik materiál odpovídá autenticitě díla na základě provedeného historického výzkumu, informacím o dobovém technickém dispozitivu a známým technickým parametrům díla?
 - Odpovídá materiál známé **historické podobě** díla na základě dostupného popisu, nebo jeho předpokladu (např. z hlediska úplnosti, odpovídajícího obsahu, varianty, jazykové verze apod.)?
 - Odpovídá materiál svým vzhledem **vizuální a zvukové identitě** díla (např. ohledně poměru stran, barevné informace apod.)?
 - Odpovídá materiál **technické identitě** díla (zejména způsobem uložení informace, např. ohledně rozsahu, formátu, snímkové frekvence, délky, počtu zvukových kanálů, datace díla vzhledem k médiu apod.)?
 - Odpovídá materiál **produkčnímu, nebo distribučnímu** formátu díla (jde o produkční master, nebo distribuční derivát vytvořený pro konkrétní situaci)?
- **Stanovení informačního obsahu materiálu.** – Jaké množství informací materiál obsahuje vzhledem k předpokládanému autentickému stavu díla (např. pokud jde o kopii novější datace obsahující více informací oproti poškozenému originálu, odpovídá množství informací očekávané datové kapacitě nosiče apod.).
- **Validace technického stavu materiálu.** – Neobsahuje médium závady, které by bránily akvizici? Např. fyzické poškození nosiče, poškození datových informací... (např. nečitelné DVD, mechanické poškození pásky, poškození souborů apod.)
- **Identifikace technické kvality a poškození.** – Jaká je technická kvalita média, jakou formou se na něm objevuje poškození, zastarávání a další vady?
- **Návrh technického zabezpečení.** – Jakou formou má být materiál na základě jeho poškození ošetřen, restaurován nebo prezervován.

Posuzované parametry materiálů se podle jednotlivých typů mohou lišit. Vycházejí z konkrétních historických rešerší jednotlivých děl, které mohou stanovit určité vlastnosti jako klíčové. Důležitým kritériem posouzení je identita materiálu vzhledem k dobovému technickému dispozitivu. Mezi obecná kritéria, která by měla být posuzována vždy, patří:

- **Délka** (vzhledem k historicky známé podobě díla)
- **Rozlišení**
- **Poměr stran**
- **Progresivní, nebo prokládané řádkování**
- **Barevné podání**
- **Televizní standard** (NTSC/PAL/SECAM...)
- **Typ média** (např. typ kazety a formát záznamu)

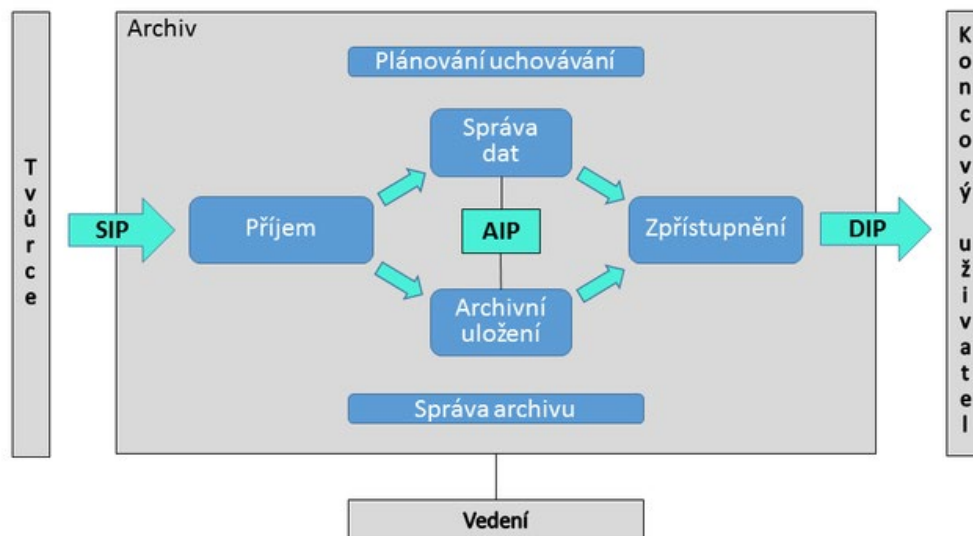
U digitálních souborů je nadále vhodné posuzovat:

- **Rozlišení** (v případě, že jde o digitalizát fyzického nosiče)
- **Poměr stran** (Aspect Ratio, např. 4:3, 16:9, Academy, Cinemascope...)
- **Bitový tok** (Bitrate, např. 10 Mbit/s, 200 kbit/s...)
- **Snímkovou frekvenci** (Frame rate, např. 23.98 fps, 24 fps, 60 fps...)
- **Velikost souboru** (File size, např. 500 MB, 10 GB...)
- **Audio kodek** (Audio codec, např. PCM, OGG, MP3...)
- **Video kodek** (Video codec, např. H.264, FFv1, Apple ProRes...)
- **Formát souboru** (File format, např. MPEG-4, Matroska, AVI, DVD-video...)
- **Počet zvukových stop** (Sound tracks, např. Mono, Stereo, 5.1...)
- **Vzorkovací frekvence** (sampling rate, např. 44.1kHz, 48kHz, 192kHz...)
- **Kompresi** (Lossy/Lossless)

Na základě posouzení uvedených charakteristik konzervátor navrhne materiály, které vzhledem k identitě díla považuje za autentické (odpovídají původní identitě díla), nebo informačně bohaté (obsahují nejvíce informací vzhledem k identitě díla).

2. ARCHIVAČNÍ FÁZE

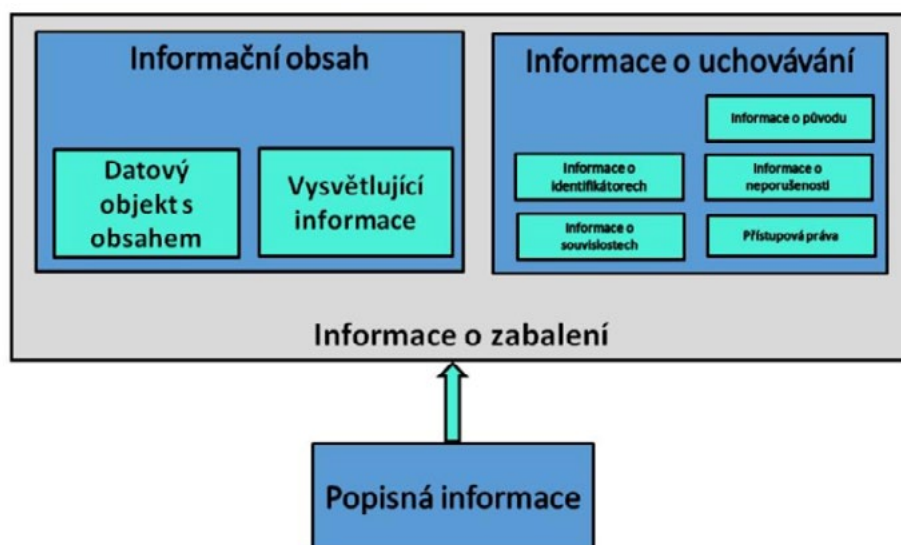
Funkční schéma systému OAIS



Účelem poloprovozu je archivní uložení digitálních objektů a jejich použití odbornou komunitou ve výzkumu a zpřístupnění veřejnosti. Tyto nároky na správnost a přesnost se realizují v repozitáři prostřednictvím signifikantních vlastností digitálních objektů (viz kapitola 1. 5.).

Realizace spočívá v detekci signifikantních vlastností repozitářem ukládaného materiálu, jejich dlouhodobém uchování a jejich prezentaci. Signifikantní vlastnosti musí být srozumitelné, interpretovatelné a vyhovovat potřebám určené cílové skupiny repozitáře. Tyto vlastnosti digitálních objektů zahrnují jejich vzhled, chování, kvalitu a použitelnost. Obecným rámcem uchování signifikantních vlastností je norma OAIS.⁶ Realizaci uchování signifikantních vlastností v balíčku AIP a jeho částech ukazuje následující schéma:

Schéma archivního balíčku AIP



⁶ ČSN ISO 14721. *Systémy pro přenos dat a informací z kosmického prostoru - Otevřený archivační informační systém - Referenční model*. 2. vyd. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014, 97 s. Třídící znak 31 9620.

Ve struktuře archivního balíčku AIP je pro uložení informací využít zejména AIP METS⁷ soubor. Ten obsahuje seznam všech digitálních objektů v AIP (originální soubory, prezervační mastery, licenční soubory, dokumentaci podání apod.), popisuje jejich vztahy a navazuje digitální objekty na jejich deskriptivní, technická, provenienční a právní metadata.⁸

2.1. Hardwarové řešení archivační fáze

Fyzická vrstva dlouhodobé archivace byla navržena dle doporučení Library of Congress a zahrnuje princip 3-2-1. Data jsou uložena ve třech identických kopiích, na dvou technologiích uložení, a alespoň jedna kopie je tzv. offline a off-site (tato kopie dat se nachází mimo areál, kde je pracoviště Videoarchivu). Jedna z technologií uložení je on-line datové úložiště s paritní ochranou dat (2x RAID-6), které zajišťuje okamžitou dostupnost dat, sekundární uložení je na magnetických datových páskách LTO generace 7 ve dvou kopiích, přičemž jedna z nich je uložena v jiném objektu pod správou Národního filmového archivu. LTO pásky jsou ukládány v elektromagneticky stíněných uzamykatelných skříních. Souborový systém LTO pásek je LTFS,⁹ který umožňuje s páskovým médiem zacházet jako s externím pevným diskem. Každá páska je označena svým kódem, který je uložen v interní databázi, a v případě potřeby přístupu k datům je třeba příslušnou pásku vyhledat v depozitáři Národního filmového archivu a vložit ji do čtecího zařízení. Po vykopírování se médium opět vrací do depozitáře. Každých 5 let je třeba datové soubory migrovat na aktuální generaci páskových médií.

2.2. Softwarové řešení archivační fáze

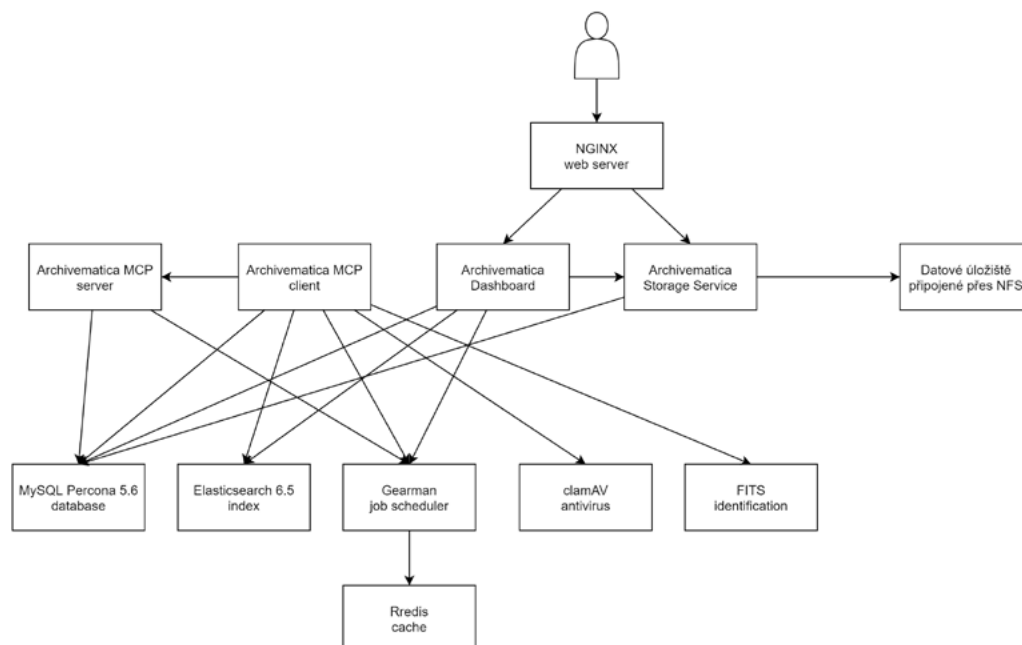
Software archivační části je taktéž založen na otevřeném řešení s cílem zajistit udržitelnost navrženého workflow tak, aby nebylo nutno např. z licenčních důvodů měnit programové komponenty během provozu. Zároveň však byly zvoleny komponenty založené na otevřených standardech a na principu tzv. microservices, kdy všechny softwarové komponenty navzájem komunikují pomocí standardizovaných protokolů a datových formátů. Případné změny v softwarové infrastruktuře bude tedy možné provádět inkrementálně bez nutnosti změny celého workflow. Hlavní softwarovou komponentou je archivační systém Archivematica¹⁰ implementující model OAIS (viz schéma na obr. č. 5). Aplikace je nasazena za pomoci technologie Docker, která umožňuje přesně replikovat instalované prostředí dle popisu jednotlivých kroků instalace. Následující diagram znázorňuje závislosti jednotlivých softwarových komponent, které vystupují jako samostatné Docker kontejnery:

7 Metadata Encoding and Transmission Standard, cit. 7. 3. 2022, <https://www.loc.gov/standards/mets/>.

8 Struktura AIP, cit. 7. 3. 2022, <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/archival-storage/aip-structure/>.

9 ISO/IEC 20919:2021, cit. 7. 3. 2022, <https://www.iso.org/standard/80598.html>.

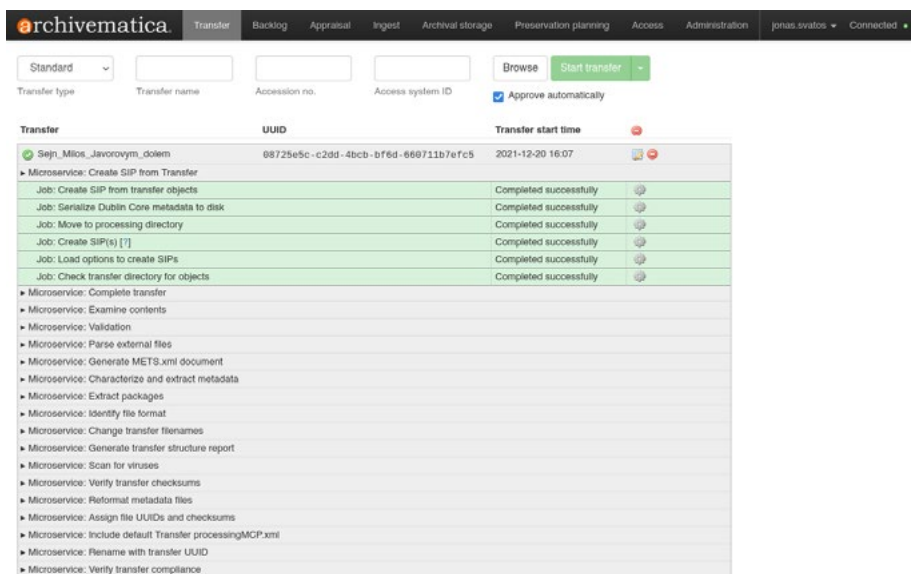
10 Systém Archivematica, cit. 7. 3. 2022, <https://www.archivematica.org/en/>.



Životní cyklus archivního balíčku je v systému Archivematica rozdělen na tři funkční části, kterými postupně data procházejí – jde o Transfer, Ingest a Archival Storage.

2.2.1. Transfer

V části Transfer probíhá výběr dat ze zdrojového úložiště pro přesun do interního pracovního úložiště systému Archivematica, který ze vstupních dat vytvoří SIP¹¹ balíček dle pravidel definovaných v konfiguraci systému. V této části taktéž probíhá párování proti přístupovému portálu AtoM na základě vložení persistentního identifikátoru do popisných metadat budoucího balíčku. Na jeho základě v další fázi probíhá nahrání DIP¹² balíčku do systému AtoM. Podrobnosti o všech funkcích v této části jsou dostupné v příslušné sekci manuálu k systému Archivematica.¹³



11 Submission Information Package.

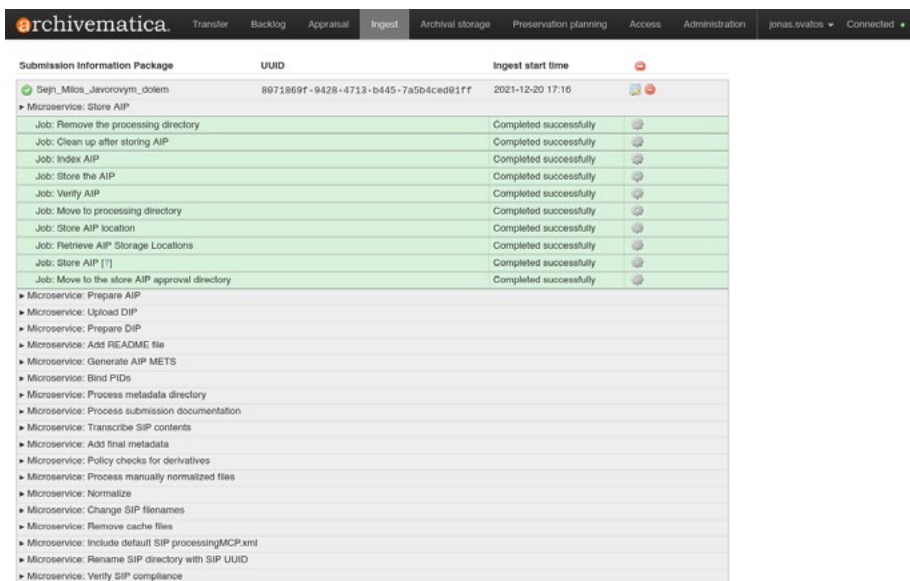
12 Dissemination Information Package.

13 Transfer, cit. 7. 3. 2022, <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/transfer/transfer/>.

2.2.2. Ingest

V Ingest části probíhají procesy vedoucí k vytvoření archivačního balíčku AIP¹⁴ a volitelně i DIP balíčku, který slouží pro zpřístupnění náhledové podoby archivovaného obsahu, z balíčku SIP, který vznikl v předchozí fázi. Proces tvorby AIP zahrnuje např. automatické pořízení technických metadat z datových souborů dle specifikace PREMIS,¹⁵ následnou normalizaci do archivačních formátů dle konfigurace archivační pipeline (tzv. transcoding), vytvoření kontrolních součtů, a výsledné zabalení výstupu do balíčku ve formátu BagIt.¹⁶ Po úspěšném provedení všech kroků dojde k uložení AIP balíčku do archivačního úložiště, v případě DIP balíčku dojde k nahrání jeho obsahu přes webovou službu do systému AtoM. Podrobnosti o všech funkcích této části jsou dostupné v příslušné sekci manuálu k systému Archivematica.¹⁷

Pracovní prostředí systému Archivematica – Ingest



Submission Information Package	UUID	Ingest start time	
Sejn_Milos_Javorovym_dotem	8671869f-9428-4713-b445-7a5b4ced011f	2021-12-20 17:16	
Microservice: Store AIP			
Job: Remove the processing directory		Completed successfully	
Job: Clean up after storing AIP		Completed successfully	
Job: Index AIP		Completed successfully	
Job: Store the AIP		Completed successfully	
Job: Verify AIP		Completed successfully	
Job: Move to processing directory		Completed successfully	
Job: Store AIP location		Completed successfully	
Job: Retrieve AIP Storage Locations		Completed successfully	
Job: Store AIP [1]		Completed successfully	
Job: Move to the store AIP approval directory		Completed successfully	
Microservice: Prepare AIP			
Microservice: Upload DIP			
Microservice: Prepare DIP			
Microservice: Add README file			
Microservice: Generate AIP METS			
Microservice: Bind PIDs			
Microservice: Process metadata directory			
Microservice: Process submission documentation			
Microservice: Transcribe SIP contents			
Microservice: Add final metadata			
Microservice: Policy checks for derivatives			
Microservice: Process manually normalized files			
Microservice: Normalize			
Microservice: Change SIP filenames			
Microservice: Remove cache files			
Microservice: Include default SIP processingMCP.xml			
Microservice: Rename SIP directory with SIP UUID			
Microservice: Verify SIP compliance			

2.2.3. Archival Storage

Část Archival Storage zajišťuje management archivačních balíčků, jejich zpětné vykopírování na pracovní úložiště, a umožňuje schvalovat či zamítnout mazání záznamů o archivačních balíčcích z databáze. Podrobnosti o všech funkcích této části jsou dostupné v příslušné sekci manuálu k systému Archivematica.¹⁸

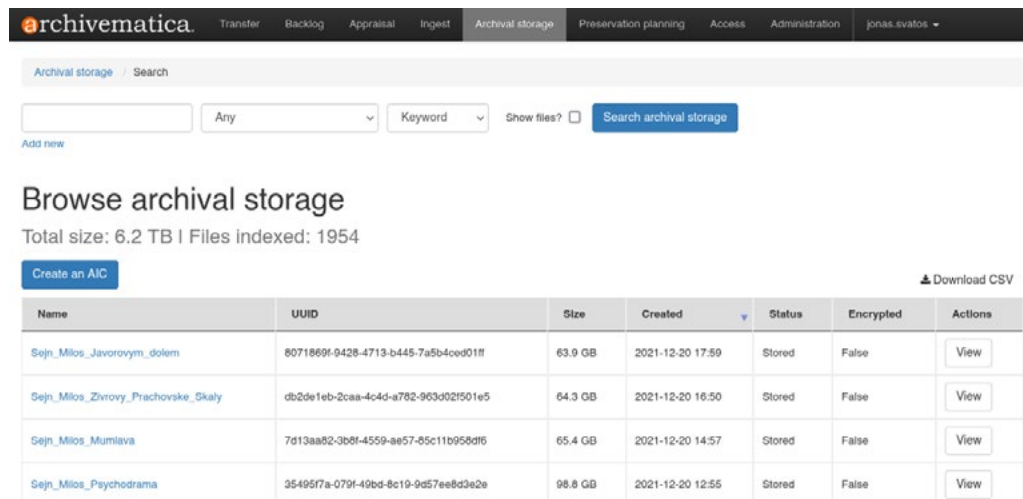
14 Archival Information Package.

15 Preservation Metadata: Implementation Strategies, cit. 7. 3. 2022, <https://www.loc.gov/standards/premis/>.

16 The BagIt File Packaging Format, RFC8493, cit. 7. 3. 2022, <https://datatracker.ietf.org/doc/html/rfc8493>.

17 Ingest, cit. 7. 3. 2022, <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/ingest/ingest/>.

18 Archival Storage, cit. 7. 3. 2022, <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/archival-storage/archival-storage/>.



archivematica Transfer Backlog Appraisal Ingest Archival storage Preservation planning Access Administration jonas.evatos

Archival storage / Search

Any Keyword Show files? Search archival storage

Add new

Browse archival storage

Total size: 6.2 TB | Files indexed: 1954

Create an AIC Download CSV

Name	UUID	Size	Created	Status	Encrypted	Actions
Sejn_Milos_Javorovym_dotem	8071869f-9428-4713-b445-7a5b4ced01ff	63.9 GB	2021-12-20 17:59	Stored	False	View
Sejn_Milos_Zivrovy_Prachovske_Skaly	db2de1eb-2caa-4c4d-a782-963d021501e5	64.3 GB	2021-12-20 16:50	Stored	False	View
Sejn_Milos_Mumstava	7d13aa82-3b8f-4559-ae57-85c11b958df6	65.4 GB	2021-12-20 14:57	Stored	False	View
Sejn_Milos_Psychodrama	35495f7a-079f-49bd-8c19-9d57ee8d3e2e	98.8 GB	2021-12-20 12:55	Stored	False	View

2.3. Technická specifikace archivačních formátů

Volba archivačních formátů byla v případě pohyblivého obrazu založena na doporučeních TC-06 (viz část 1. 2.):

Zdrojový master:

- V případě digitalizace z fyzického nosiče:
 - Kontejner: Matroska
 - Video kodek: FFv1¹⁹
 - Audio kodek: PCM 24-bit, 48kHz, signed Little-Endian
- V případě nativně-digitálního díla se archivuje nejlepší dostupná verze dle signifikantních vlastností (viz část 1. 5.)

Archivační master:

- Kontejner: Matroska
- Video kodek: FFv1²⁰
- Audio kodek: PCM 24-bit, 48kHz, signed Little-Endian

Autorizovaná reprezentace:

- Kontejner: Matroska
- Video kodek: FFv1
- Audio kodek: PCM 24-bit, 48kHz, signed Little-Endian

Archivní AV formáty:

- FFv1/PCM v kontejneru Matroska
- Apple ProRes/PCM v kontejneru Apple QuickTime
- H.264/AAC v kontejneru MP4

Archivní textové a ostatní datové formáty: dle doporučení Library of Congress²¹

¹⁹ Specifikace viz FFV1 Video Coding Format Versions 0, 1, and 3 RFC 9043, cit. 7. 3. 2022, <https://datatracker.ietf.org/doc/rfc9043/>.

²⁰ Specifikace viz FFV1 Video Coding Format Versions 0, 1, and 3, cit. 7. 3. 2022, <https://datatracker.ietf.org/doc/rfc9043/>.

²¹ Library of Congress Recommended Formats Statement, cit. 7. 3. 2022, <https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022.pdf>.

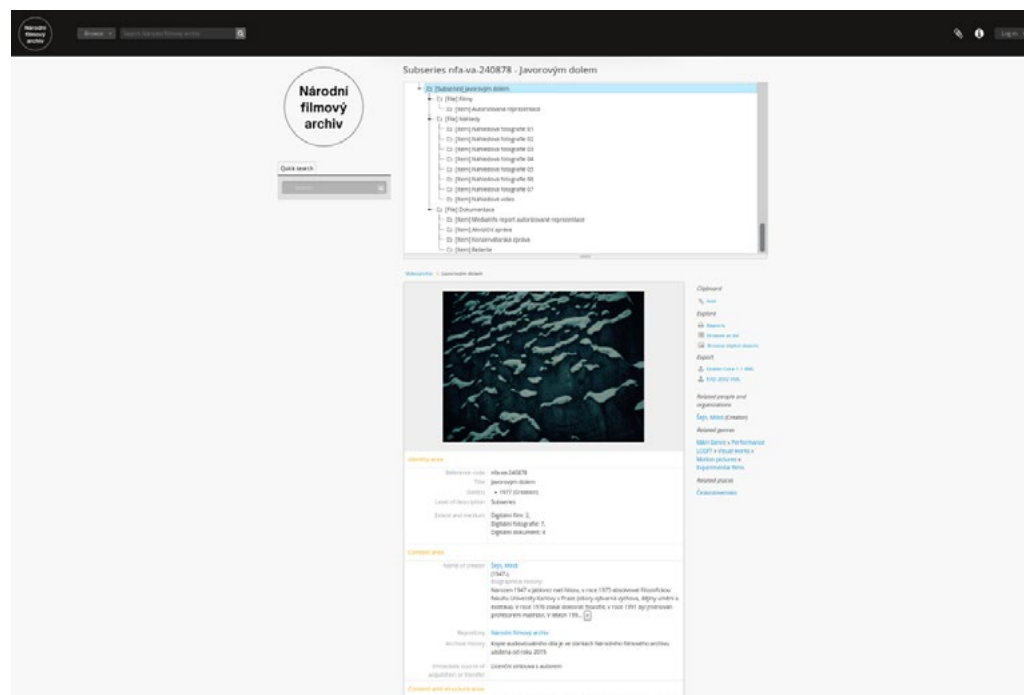
3. PREZentační fáze

3.1. Popis přístupového portálu

Katalog je provozován nad open-source systémem AtoM (Access to Memory), který je vyvíjen společností Artefactual jakožto komplementární řešení k archivnímu systému Archivematica. Propojení těchto systémů je popsáno ve veřejně dostupné dokumentaci²² a je založeno na přenesení informací ze systému Archivematica ve formě SIP balíčku, který obsahuje jak popisné informace ve formátu PREMIS, tak i náhledové verze audiovizuálních děl.

Katalog, dostupný formou webového přístupu na adrese <https://videoarchiv.nfa.cz/katalog>, má veřejně přístupnou část pro nepřihlášené uživatele, ve které jsou vidět popisné informace a náhledové fotografie pro veškerá zveřejněná díla. Část pro přihlášené uživatele jim umožňuje zhlédnout i náhledové verze samotných archivovaných děl ve webovém videopřehrávači. Více o funkčnosti přístupového portálu je k dispozici v manuálu k použití dostupném na webových stránkách projektu.²³

Pracovní prostředí
Přístupového portálu AtoM



3.2. Hardwarové řešení přístupového portálu

Portál AtoM sdílí hardware s archivním systémem Archivematica ve formě odděleného virtuálního serveru na serveru pro aplikace.

22 Dokumentace k aktuální verzi AtoM viz AtoM documentation, cit. 7. 3. 2022, <https://www.accesstomemory.org/en/docs/2.6/>.

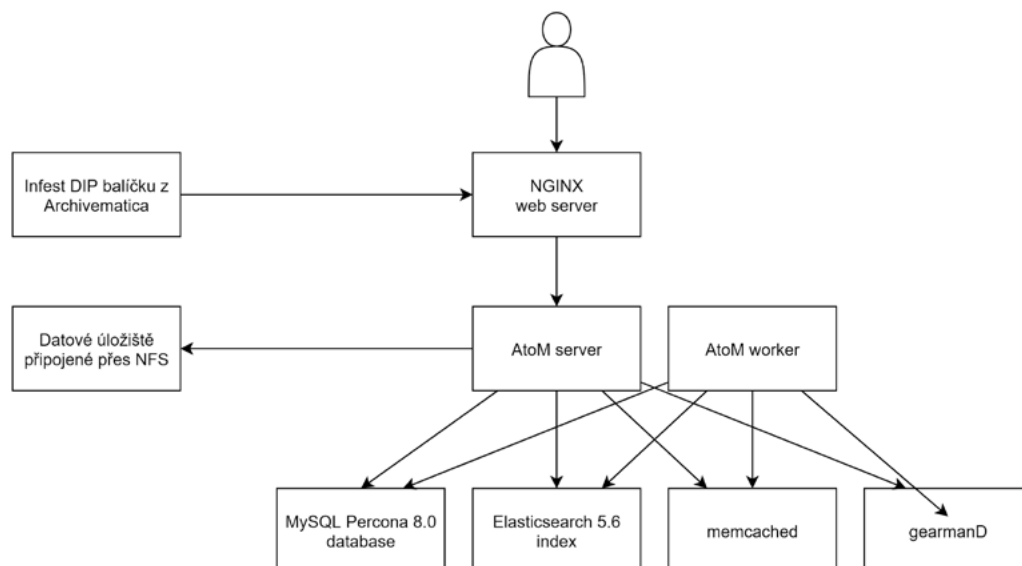
23 Webová prezentace projektu viz Videoarchiv, cit. 7. 3. 2022, <https://videoarchiv.nfa.cz>.

3.3. Softwarové řešení přístupového portálu

Přístupový portál je implementován za pomoci systému Access to Memory²⁴ (dále AtoM). Přidělování perzistentních identifikátorů zajišťuje interní systém Národního filmového archivu nazvaný ID-system, na který je editační rozhraní databáze napojeno.

AtoM je na serveru instalován dle pokynů v dokumentaci projektu.²⁵ Následující funkční schéma znázorňuje propojení jednotlivých Docker kontejnerů a napojení na externí systémy:

Logické schéma propojení softwarových komponent prezentačního systému



Pro potřebu Národního filmového archivu bylo třeba provést několik úprav ve zdrojovém kódu, aby funkcionální vyhovovala požadavkům projektu. Jednalo se zejména o:

- Implementaci generování unikátních identifikátorů skrze API aplikace “ID-system”, která byla vyvinuta v Národním filmovém archivu za účelem sjednocení generování identifikátorů (taktéž označována jako Resolver).
- Úprava nastavení LDAP konektoru pro integraci přihlašování pomocí centrální správy účtů uživatele výsledku.
- Úprava grafické šablony dle vizuálního stylu uživatele výsledku.

Upravený systém, založený na původním zdrojovém kódu verze AtoM v2.6, je dostupný v Git repozitáři²⁶ uživatele výsledku.

24 Library of Congress Recommended Formats Statement – 2020–2021, cit. 7. 3. 2022, <https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022.pdf>.

25 Dokumentace instalace viz Docker Compose, cit. 7. 3. 2022, <https://www.accesstomemory.org/en/docs/2.6/dev-manual/env/compose/>.

26 Zdrojový kód je dostupný na adrese: <https://git.digilab.nfa.cz/naki-videoarchiv/atom>.

4. EKONOMICKÉ PARAMETRY POLOPROVOZU

Poloprovoz předpokládá testování automatizovaného systému ve zkušebních podmínkách a s nižší zátěží, aby bylo zjištěno, jaké kapacity provozu je schopen. Je však rozdíl mezi automatizovanými systémy výrobními a systémy jiného určení. Zejména u poskytování služeb a v oblasti školství a společenských věd nelze automatizované systémy hodnotit jejich produkčními parametry, tento provoz je často historicko-výzkumného charakteru, takže není srovnatelný s žádnou tržním způsobem poskytovanou službou. Jde o konzervátorské zpracování vyžadující specifickou kvalifikaci, dostupnou pouze v akademických a paměťových institucích.

Archivační řetězec lze ekonomicky využívat několika různými způsoby:

- Poplatky za jednotlivé úkony archivace, zpracování jednotlivých balíčků, jednotlivých nosičů apod.
- Poplatky rámcové nebo paušální s definovaným objemem zpracování za časové období (např. roční).
- Jednotlivé poplatky nebo rámcové poplatky mohou být závislé na dalších smluvních podmínkách ve zvoleném obchodním modelu, kdy je uzavírání licence na užití autorského díla, resp. depozitní smlouva pro archivaci může být podmíněna opcí na uzavření licenční smlouvy.
- Vzhledem k předmětu archivace, kterým je umění jako kulturní dědictví na nerozvinutém trhu, je třeba brát také v úvahu, že nebude možné najít udržitelný obchodní model soukromoprávního charakteru a provoz bude vyžadovat veřejnou podporu.

Ekonomické parametry závisí tedy na zvolené sbírkové politice, resp. záleží na obchodním modelu, ve kterém je repozitář provozován. Výsledek je také pouze systémem zpracování archivního materiálu, kdy náklady na provoz kompletního repozitáře představují rovněž náklady na dlouhodobé uložení, které mohou přesahovat náklady na zpracování, tedy prvotní přijetí do repozitáře.

Způsob stanovení právního a ekonomického rámce provozu archivního systému a jeho využívání původci archiválií popisuje Technicko-organizační směrnice provozu archivního systému pohyblivého obrazu.²⁷ To odpovídá požadavkům norem implementace repozitářů na stanovení procesů a zodpovědnosti.

²⁷ Technicko-organizační směrnice provozu archivního systému pohyblivého obrazu, cit. 7. 3. 2022, https://www.mkcr.cz/doc/cms_library/hneleg_smernice_technicko_organizacni_smernice_dg20p020vv025-14167.pdf.

5. PŘÍLOHY

5.1. Vzor konzervátorské zprávy

Konzervátorská zpráva

Název díla, datace: Bez názvu, 1998

Jméno a příjmení autora: Jan Novák

Vyhotovil/a: Digitální operátor, Národní filmový archiv

Datum vyhotovení: 21. 7. 2021

Poznámky k produkci díla

Typ nosiče	born-digital video
Zdrojový master	
Fyzický popis nosiče (vč. identifikátorů, štítků apod.)	N/A
Název souboru	
Fotodokumentace zdrojového nosiče	N/A
Timecode díla na nosiči	00:00:00:00-00:05:37:09
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Video kodek	ProRes 422 HQ
Audio kodek	PCM Little / Signed, 48 kHz, 16 bits
Poměr stran obrazu	4:3
Počet audio stop	2
Byla na zdrojovém nosiči identifikována nějaká poškození? Pokud ano, jaká? Došlo k nějakým úpravám nosiče?	N/A
Parametry digitalizace	
Přehrávací zařízení	N/A
Digitalizační zařízení	N/A
Bylo přistoupeno ke korekcím signálu zdrojového nosiče (např. zvýšení citlivosti snímače, či upravený odstup signálu od šumu apod.)? Pokud ano, k jakým?	ne
Popis softwarového řetězce	nfa-transcoder, profil „lossless“
Bylo přistoupeno k úpravám v digitálním prostředí? Pokud ano, k jakým?	ne
Jaký byl důvod výběru tohoto zdrojového masteru?	jedná se o master dodaný autorem
Archivní master	
Název souboru	AM-Bez_nazvu-01.mov
Délka	5 min 37 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	Matroska
Video kodek	FFv1
Audio kodek	FLAC
Poměr stran obrazu	4:3
Počet audio stop	2

Autorizovaná reprezentace	
Název souboru	AR-Bez_nazvu-01.mov
Délka	5 min 37 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Video kodek	ProRes 422 HQ
Audio kodek	PCM Little / Signed, 48 kHz, 16 bits
Poměr stran obrazu	4:3
Počet audio stop	2

Typ nosiče	born-digital video
Zdrojový master	
Fyzický popis nosiče (vč. identifikátorů, štítků apod.)	N/A
Název souboru	2_Bez_nazvu.mov
Fotodokumentace zdrojového nosiče	N/A
Timecode díla na nosiči	00:00:00:00-00:06:08:08
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Video kodek	ProRes 422 HQ
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0
Byla na zdrojovém nosiči identifikována nějaká poškození? Pokud ano, jaká? Došlo k nějakým úpravám nosiče?	N/A
Parametry digitalizace	
Přehrávací zařízení	N/A
Digitalizační zařízení	N/A
Bylo přistoupeno ke korekcím signálu zdrojového nosiče (např. zvýšení citlivosti snímače, či upravený odstup signálu od šumu apod.)? Pokud ano, k jakým?	ne
Popis softwarového řetězce	nfa-transcoder, profil „lossless“
Bylo přistoupeno k úpravám v digitálním prostředí? Pokud ano, k jakým?	ne
Jaký byl důvod výběru tohoto zdrojového masteru?	jedná se o master dodaný autorem
Archivní master	
Název souboru	AM-Bez_nazvu-02.mov
Délka	6 min 8 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	Matroska
Video kodek	FFv1
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0

Autorizovaná reprezentace	
Název souboru	AR-Bez_nazvu-02.mov
Délka	6 min 8 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Video kodek	ProRes 422 HQ
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0

Typ nosiče	born-digital video
Zdrojový master	
Fyzický popis nosiče (vč. identifikátorů, štítků apod.)	N/A
Název souboru	3_Bez_nazvu.mov
Fotodokumentace zdrojového nosiče	N/A
Timecode díla na nosiči	00:00:00:00-00:03:15:00
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Video kodek	ProRes 422 HQ
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0
Byla na zdrojovém nosiči identifikována nějaká poškození? Pokud ano, jaká? Došlo k nějakým úpravám nosiče?	N/A
Parametry digitalizace	
Přehrávací zařízení	N/A
Digitalizační zařízení	N/A
Bylo přistoupeno ke korekcím signálu zdrojového nosiče (např. zvýšení citlivosti snímače, či upravený odstup signálu od šumu apod.)? Pokud ano, k jakým?	ne
Popis softwarového řetězce	nfa-transcoder, profil „lossless“
Bylo přistoupeno k úpravám v digitálním prostředí? Pokud ano, k jakým?	ne
Jaký byl důvod výběru tohoto zdrojového masteru?	jedná se o master dodaný autorem
Archivní master	
Název souboru	AM-Bez_nazvu-03.mov
Délka	3 min 15 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	Matroska
Video kodek	FFv1
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0

Autorizovaná reprezentace	
Název souboru	AR-Bez_nazvu-03.mov
Délka	3 min 15 s
Obrazové rozlišení	720x576
Snímková frekvence	25 fps
Souborový formát	MPEG-4
Vídeo kodek	ProRes 422 HQ
Audio kodek	N/A
Poměr stran obrazu	4:3
Počet audio stop	0

5.2. Soupis technického vybavení

Počet	Položka	Pořízený produkt	Konfigurace
2	Pracovní stanice	Supermicro SuperWorkstation	OS Rocky Linux 8.4 2x CPU Intel Xeon Silver 4208 64GB RAM 512GB NVMe systémový disk GPU NVidia Quadro RTX 4000 8TB pro data v RAID1 10Gbit SFP+ Ethernet síťové rozhraní
2	Koloristický monitor	Dell UP3221Q	Rozlišení UHD (3996x2160) Hloubka zobrazení 10-bit/kanál Rozhraní 2x DisplayPort, 2x HDMI HW funkce nastavení LUT Integrovaná kalibrační sonda Stínidlo ambientního světla
2	Digitalizační karta	Blackmagic DeckLink Studio 4K	Maximální rozlišení 4K Maximální rozlišení HD Výstup HDMI 4K Výstup 6G SDI Výstup YUV Výstup Composite Výstup S-Video Vstupy HDMI 4K Vstupy 6G SDI Vstupy YUV Vstupy Composite Vstupy S-Video Připojení PCI Express 2.0 x4
1	CRT Monitor	Sony PVM-14L2	
1	Osciloskop	Tektronix 1731	
1	VHS/S-VHS přehrávač, norma PAL	Sony SVP-5600P	podporované kazetové formáty: VHS, S-VHS podporované standardy: PAL obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	VHS/S-VHS přehrávač, norma NTSC	Sony SVP-5600	podporované kazetové formáty: VHS, S-VHS podporované standardy: NTSC obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	Digital Betacam/ Betacam SP přehrávač norma PAL	Sony DVW-A510P	podporované kazetové formáty: Digital Betacam, Betacam SP podporované standardy: PAL, NTSC obrazový a zvukový výstup: SDI napájení: 230 V / 50 Hz
1	Digital Betacam/ Betacam SP přehrávač norma NTSC	Sony J-30SDI	podporované kazetové formáty: Digital Betacam, Betacam SP podporované standardy: PAL obrazový a zvukový výstup: SDI napájení: 230 V / 50 Hz
1	DVCAM/MiniDV přehrávač, norma PAL	DSR-1500AP	podporované kazetové formáty: DVCAM, miniDV podporované standardy: PAL obrazový a zvukový výstup: SDI, FireWire napájení: 230 V / 50 Hz
1	DVCAM/MiniDV přehrávač, norma NTSC	DSR-1500A	podporované kazetové formáty: DVCAM, miniDV podporované standardy: NTSC obrazový a zvukový výstup: SDI, FireWire napájení: 230 V / 50 Hz

1	U-Matic přehrávač, norma PAL	Sony VO-9850P	podporované kazetové formáty: U-Matic, U-Matic SP podporované standardy: PAL obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	U-Matic přehrávač, norma NTSC	Sony VO-9850	podporované kazetové formáty: U-Matic, U-Matic SP podporované standardy: NTSC obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	Hi-8/Video8/Digital8 přehrávač, norma PAL	Sony GV-D800E	podporované kazetové formáty: Hi-8/Video8/Digital8 podporované standardy: PAL obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	Hi-8/Video8/ Digital8 přehrávač, norma NTSC	Sony GV-D800	podporované kazetové formáty: Hi-8/Video8/Digital8 podporované standardy: NTSC obrazový a zvukový výstup: Kompozitní napájení: 230 V / 50 Hz
1	Archivní úložiště	Supermicro SuperStorage	OS oVirt Node 4.4 2x CPU Intel Xeon Silver 4208 32GB RAM 512GB SATA SSD systémový disk 12x 8TB SAS 7.2k HDD SAS3 RAID controller 2x RAID5 array s kapacitou 40TB 10Gbit SFP+ Ethernet síťové rozhraní SAS3 rozhraní pro připojení LTO páskové mechaniky
1	Server pro aplikace	Supermicro SuperServer	OS oVirt Node 4.4 2x CPU Intel Xeon Silver 4208 128GB RAM 512GB SATA SSD systémový disk 10Gbit SFP+ Ethernet síťové rozhraní
1	Síťový přepínač, rozhraní SFP+, 10Gbit	Cisco SX550X-12G	12 SFP+ portů Web-management Propustnost min. 240Gbit/s Provedení do racku - 1U
1	Síťový přepínač, rozhraní Ethernet, 1Gbit	NETGEAR ProSafe S3300-28X	28 Ethernet portů, 2 SFP+ porty Web-management Propustnost min. 36Gbit/s Provedení do racku - 1U
1	LTO7 pásková mechanika, rozhraní SAS3	Quantum LTO-7 HH SAS external Drive	komunikační rozhraní: 6Gb SAS napájení: 230 V / 50 Hz Tabletop provedení.
1	Přenosná pracovní stanice	Lenovo ThinkPad X390	Intel Core i7 8565U Whiskey Lake dotykový 13.3" IPS antireflexní 1920 × 1080 RAM 16GB DDR4 Intel UHD Graphics 620 SSD 512GB podsvícená klávesnice webkamera USB 3.2 Gen 1, USB-C, čtečka otisků prstů WiFi 5 48 Wh baterie