



MANUSCRIPTORIUM

Diverzifikace částí a funkcionalit komplexního digitálního dokumentu

www.manuscriptorium.com



1	Úvod	3
2	Možnosti využití IIIF v distribuovaném prostředí	4
3	Ověřovací instalace	5
3.1	Dostupnost výsledného řešení.....	5
3.2	Technický popis implementovaného řešení	7
3.2.1	Loris2	7
3.2.2	Globální obrazový repositář	7
3.2.3	Budoucí rozsah centralizace obrazových dat	8
3.2.4	Optimalizace přenosových rychlostí v běžném (polo)provozu	8
3.3	Implementace IIIF Presentation API v prostředí distribuovaných digitálních repositářů	8
3.3.1	IIIF Presentation API při poskytování přístupu k agregovanému obsahu	9
3.3.2	Využití IIIF Presentation API při přístupu k strukturálním metadatům distribuovaných repositářů kompatibilních s IIIF	11
4	Ukázky využití IIIF v prostředí Manuscriptoria	12
5	Závěr	14



1 Úvod

Cílem našich výzkumných a vývojových aktivit je usnadnění přístupu k výsledkům celosvětové digitalizace historických knižních fondů. Každý digitalizační projekt dosud produkuje a zpřístupňuje digitální obsah (data i metadata) v různých formátech, kvalitách, strukturách a pomocí různých nástrojů.

V zájmu zvýšení využitelnosti výsledků digitalizace („otevírání archivů“) iniciativa IIF definuje pravidla přístupu k obrazovým repositářům: specializované sady pravidel předepisují producentům digitálního obsahu, jak zpřístupňovat informace o sbírkách, dokumentech i jednotlivých stránkách/obzích standardizovaně prostřednictvím aplikačních rozhraní, ke kterým jsou zároveň v komunitě IIF vyvíjeny dedikované aplikace (včetně zobrazovacích aplikací pro koncové uživatele).

Zvyšování interoperability jednotlivých repositářů právě adaptací IIF je jistě žádoucí, ale dlouhotrvající proces - „otevírání archivů“ bude pozvolné a z hlediska současných uživatelů nepřijatelně pomalé.

Využili jsme toho, že Manuscriptorium agreguje ve velkém měřítku produkci digitalizačních pracovišť z celého světa a hledali jsme způsob, jak jeho prostřednictvím plošně aplikovat IIF protokol v nesourodém prostředí digitalizovaných historických knižních fondů, čímž bychom dosáhli dalšího stupně homogenizace přístupu k obrazovým informacím v dané oblasti zájmu.

Výsledkem prací je inovované Manuscriptorium, které v reálném čase dokáže zprostředkovat přístup k obrazům libovolného partnera standardizovaným způsobem. Přístupovat k tomuto původně roztržitému obsahu pak lze nejen v uživatelském prostředí Manuscriptoria, ale i prostřednictvím libovolného jiného nástroje, který podporuje příslušná IIF API. To okamžitě přináší výrazně lepší interoperabilitu v prostředí digitalizovaných historických fondů a zvyšuje využitelnost produkovaného digitálního obsahu napříč různými komunitami koncových uživatelů (například EUROPANA a jiní alternativní zpřístupňovatelé obsahu budou tuto inovaci využívat ve prospěch svých uživatelů).

Konkrétně Manuscriptorium s podporou IIF pro agregované dokumenty umožňuje například:

- Koncovému uživateli persistentně adresovat libovolný obraz nebo i jeho libovolný výřez (student, pedagog, badatel ve své práci mohou persistentně odkazovat URL odkazem na konkrétní iluminovanou iniciálu libovolného dokumentu v libovolném repositáři agregovaném Manuscriptorem bez ohledu na současné či budoucí vlastnosti daného repositáře).
- Koncovému uživateli umožnit otevřít libovolný komplexní digitální dokument v libovolné SW aplikaci s podporou IIF a projektům třetích stran umožnit načíst celé kolekce a jejich komplexní digitální dokumenty a pracovat s nimi ve vlastním SW/interface¹.

¹ V srpnu 2017 přechází tato funkcionalita z režimu poloprovozu do plného provozu a k datu vytvoření textu je dostupná pro vybrané kolekce zveřejňované postupně v rámci IIF Universe (<https://graph.global/static/data/universes/iif/universe.json>); tento obsah bude postupně rozšiřován, aby pokryl kompletní kolekce prezentované Manuscriptorem

Při plnění vytyčených úkolů jsme se připojili ke komunitě IIF a čerpali ze sdíleného know-how. Instalovali jsme jeden z doporučených image serverů (Loris2) a analyzovali možnosti jeho modifikace pro potřeby distribuovaného prostředí. Došli jsme k závěru, že Loris2 lze efektivně kustomizovat tak, aby mechanismy přístupu k obrazům přes HTTP protokol byly optimálně využitelné při zprostředkování přístupu k distribuovanému obrazovému obsahu.

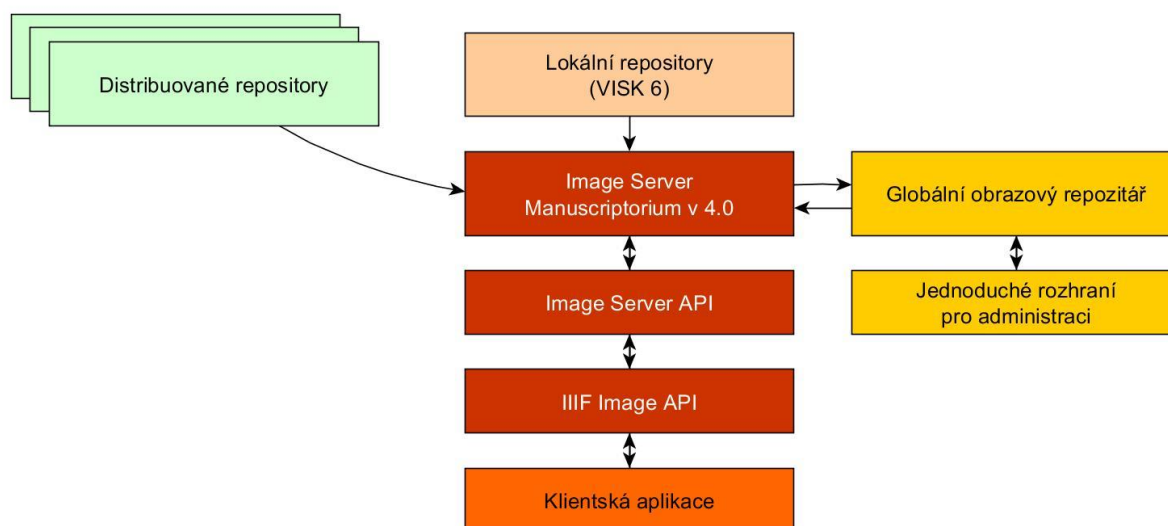
Výsledkem implementace a kustomizace byl nejprve poloprovaz a nyní již plný provoz Image Serveru Manuscriptoria, včetně modifikovaného koncového uživatelského prostředí Manuscriptoria a nástroje se podařilo implementovat i do stávající verze Manuscriptoria (verze 3).

2 Možnosti využití IIF v distribuovaném prostředí

Protokol IIF předepisuje standard, jak má vypadat komunikace mezi klientem a obrazový repositář. My jsme hledali způsob, jak nasadit IIF tam, kde je prostřednictvím protokolu komunikováno s *prostředníkem* mezi klientem a mnoha cílovými obrazovými repositáři.

Využili jsme existující image server Loris2, který je pro podobné účely vhodný, protože již obsahuje implementovaný modul HTTP resolver a po nutných kustomizačních zásazích potřebám našeho záměru perfektně vyhovuje.

Smyslem kustomizačních zásahů je přeměnit původní provozní cache serveru, v přehledně organizovaný globální repositář, kde je pro potřeby prezentace a zajištění rychlého a persistentního přístupu ukládán agregovaný obrazový obsah relativně trvale a lze jej lze poměrně snadno spravovat.



Obrázek 1: zjednodušené schéma řešení implementace IIF compliant Image Serveru do Manuscriptoria

Princip řešení spočívá v tom, že původně plně distribuované prostředí Manuscriptoria se stává systémem s částečnou centralizací prezentovaného obsahu.

Pro potřeby poloprovozu a později plného provozu byl připraven webový server s dostatečným výkonem a přiměřeně velkým datovým prostorem. Tento prostor byl v počáteční fázi zcela prázdný.

Systém funguje tak, že jakýkoliv požadavek koncového uživatele na zobrazení detailních obrazů historických dokumentů zahájí nejprve komunikaci mezi Image Serverem Manuscriptoria a vzdáleným obrazovým repositářem, který předá do Manuscriptoria obraz nejvyšší dostupné kvality.

Takto získaný obraz je podle definovaných pravidel umístěn v Globálním obrazovém repositáři (GOR) Manuscriptoria, které je organizováno s využitím metadat o spolupracujících repositářích a metadat komplexních digitálních dokumentů.

Následně je Image Server Manuscriptoria použit k tomu, aby prezentoval obraz z GOR v požadované kvalitě klientské aplikaci prostřednictvím aplikačního rozhraní IIIF Image API.

Toto řešení má následující výhody:

- 1) je funkční od prvního okamžiku implementace a spuštění (odpadá potřeba úvodní fáze plnění daty),
- 2) snižuje závislost služeb Manuscriptoria na dostupnosti partnerských obrazových repositářů,
- 3) shromažďuje pouze data těch zdrojů, které jsou skutečně koncovými uživateli používány (lepší efektivita při využívání dedikovaného prostoru pro data)².

3 Ověřovací instalace

3.1 Dostupnost výsledného řešení

Popsané řešení jsme implementovali v rámci poloprovozu Manuscriptoria a nově je též všem uživatelům dostupné v rámci současného Manuscriptoria na www.manuscriptorium.com (nebo přímo www.manuscriptorium.com/apps).

Konkrétní odkazy na ukázky prostřednictvím komunikačního protokolu IIIF Image API:

- IIIF-compliant Image Server v prostředí distribuovaných digitálních úložišť je dostupný na <http://images.manuscriptorium.com/loris/> a lze s ním komunikovat dle pravidel [IIIF Image API](#)

² Výjimkou jsou velmi rozměrné obrazy – proto jsme pro vybrané vzdálené repositáře umožnili v rámci administrace spouštět na pozadí stahování obsahu předem.



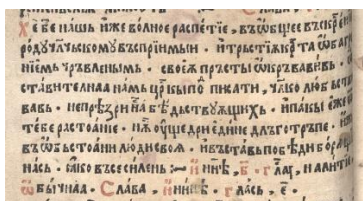
Příklady přímých požadavků:

- Repositář Národní knihovny České republiky:
dokument *Horae beatae Mariae virginis*; XXIII.F.197; Národní knihovna České republiky;
Praha; Česko:
 - http://images.manuscriptorium.com/loris/AIPDIG-NKCR_XXIII_F_197_076XW55-cs/ID0027r/info.json
Interpretace: žádost o manifest konkrétního obrázku
 - http://images.manuscriptorium.com/loris/AIPDIG-NKCR_XXIII_F_197_076XW55-cs/ID0027r/full/full/0/default.jpg
Interpretace: žádost na základě manifestu daného obrázku o kvalitu default
 - http://images.manuscriptorium.com/loris/AIPDIG-NKCR_XXIII_F_197_076XW55-cs/ID0027r/350,500,550,450/full/0/default.jpg
Interpretace: žádost o určitý výřez daného obrázku v kvalitě default:



Obrázek 2: ukázka odezvy konkrétního IIIF požadavku

- Repositář Rumunské národní knihovny:
dokument *Triod Postnaja*, izd. *verojatno diakonom Koresi*; CR XVI IV 3; National Library of Romania; Bucharest; Romania:
 - http://images.manuscriptorium.com/loris/NLR_-NLR_CR_XVI_IV_3_1VZOR98-ru/SURF_CR_XVI_IV_3_ID0015/info.json
Interpretace: žádost o manifest konkrétního obrázku
 - http://images.manuscriptorium.com/loris/NLR_-NLR_CR_XVI_IV_3_1VZOR98-ru/SURF_CR_XVI_IV_3_ID0015/full/full/0/default.jpg
Interpretace: žádost na základě manifestu daného obrázku o kvalitu default
 - http://images.manuscriptorium.com/loris/NLR_-NLR_CR_XVI_IV_3_1VZOR98-ru/SURF_CR_XVI_IV_3_ID0015/400,600,870,470/full/0/default.jpg
Interpretace: žádost o určitý výřez daného obrázku v kvalitě default:



Obrázek 3: ukázka odezvy konkrétního IIIF požadavku



3.2 Technický popis implementovaného řešení

3.2.1 Loris2

Loris2 je open-source image server, který podporuje komunikaci pomocí protokolu IIIF. Zdrojový kód Loris2 je napsán v jazyce Python. Tento Image Server lze nainstalovat pouze v OS LINUX. Naše ověřovací instalace proběhla v systému Ubuntu/Debian. Na serveru běží webserver Apache2 s modulem WSGI a je vyžadována a knihovna Python verze 2.7 (NIKOLIV 3) s modulem PIP a Pillow.

Loris2 je poměrně versatlní, protože pro manipulaci s obrazy může využít knihovnu OpenJPEG, ale i Kakadu. Bude-li to v budoucnu potřeba, lze systém rozšířit o podporu dalších formátů obrazů s možností zobrazení ve vlastním (rozuměj Manuscriptoriu vlastním, případně komunitě IIIF vlastním či jiném) nativním klientském prohlížeči. V současnosti však k zavedení takové podpory nevidíme důvod.

Loris2 byl námi dále rozšířen o podporu transferu a modifikace EXIF a XMP údajů díky propojení s knihovnou PYEXIV2, kteroužto funkci považujeme za přínosnou – umožní například vkládat informace o copyrightu a licencích přímo do vnitřností prezentovaného obrazového souboru.

Klíčovou součástí řešení je modifikovaný HTTP Resolver, který se stará o přístup ke vzdáleným obrazům a jejich uložení v Globálním obrazovém repozitáři k dalšímu použití. Tento resolver organizuje v úložišti data pomocí evidovaných informací o spolupracujícím repozitáři a metadat získaných z každého agregovaného komplexního digitálního dokumentu. Stručně vyjádřeno: jednotlivé obrazy jsou umísťovány do adresářové struktury podle toho, z jakého profilu pocházejí a kde jsou uloženy (resp. byly uloženy v okamžiku digitalizace)³.

3.2.2 Globální obrazový repozitář

Jak již bylo vysvětleno výše, systém využívá organizované centrální datové úložiště. Systém využívá existence strukturálních metadat agregovaných dokumentů, aby obsah účelně organizoval.

Součástí standardní žádosti o obrazová data je pouze identifikace dokumentu a identifikace příslušného obrazu. To k efektivní správě centralizovaných dat nestačí, a proto systém na základě těchto informací dotazuje databázi metadat agregovaných dokumentů a zjišťuje další informace, jako například zdroj digitalizovaného dokumentu či identifikátor místa uložení (knihovny).

Obrazová data jednoho dokumentu je pak možno umístit v adresářové struktuře organizované takto:

- root centrálního datového prostoru
 - identifikace místa uložení exempláře⁴
 - rozlišovací kód odvozený z identifikace exempláře
 - identifikace exempláře
 - identifikace profilu přispěvatele

³ Tato problematika velmi úzce souvisí s problematikou indexu historických fondů a digitálními konkordancemi, jak je popsána a do režimu poloprovozu uváděna v roce 2016. Předpokládáme další vývoj směrem k propojení obou systémů v rámci poloprovozu v roce 2017.

⁴ Systém pro evidenci a identifikaci míst uložení platný v celosvětovém měřítku je též jednou ze součástí vývojových a výzkumných aktivit souvisejících s Manuscriptoriem.

- obrazová data konkrétního digitálního dokumentu

3.2.3 Budoucí rozsah centralizace obrazových dat

Protože design systému snižuje závislost agregátora na partnerských repositářích, předpokládáme, že osvědčí-li se řešení při plném provozu, bude systém rozšířen tak, aby v době nižší zátěže automaticky stahovat kompletní obrazová data všech používaných dokumentů.

Uvažujeme takové řešení, že systém vyhodnotí četnost využívání záznamů o dokumentech, prověří obsah GOR a pokud jejich data nebudou dosud centralizována, zahájí jejich stahování prostřednictvím odesílání příslušných žádostí Image Serveru Manuscriptoria.

3.2.4 Optimalizace přenosových rychlostí v běžném (polo)provozu

Součástí provedených činností byla i simulace zátěže. Testy potvrdily, že s doporučenou konfigurací bude možno realizovat nejen poloprovaz, ale později i přepnutí do režimu plného provozu.

Během testů se potvrdilo očekávání, že je vhodné speciálně ošetřit poskytování náhledů na obsah digitální kopie, které je v Manuscriptoriu řešeno formou rozsáhlejších galerií. S ohledem na optimalizaci doby vyřízení požadavku klienta, nejsou tedy náhledové obrázky (thumbnails) vždy generovány přímo Image Serverem Manuscriptoria, ale v případě, že partnerský repositář již nabízí vhodnou náhledovou kvalitu, je tato přímo použita. Tzn. kustomizované API rozhraní Loris2 je nyní připraveno rozlišit požadavek na detailní obraz od požadavku na náhledové obrázky. Požadované thumbnaily pak přímo přenáší z partnerského repositáře, pokud tam již existují a ukládá je v nezměněné podobě v dedikované části GOR, odkud jsou přenášeny přímo klientské aplikaci.

Detailní (objemné) obrazy jsou přenášeny pouze tehdy, pokud o ně uživatel nebo klientská aplikace explicitně požádá při zobrazení detailu či výřezu obrazu, případně při automatické agregaci obrazového obsahu na pozadí.

3.3 Implementace IIIF Presentation API v prostředí distribuovaných digitálních repositářů

IIIF Presentation API (<http://iiif.io/api/presentation/2.1/>) zprostředkovává informace o existenci dokumentů a celých kolekcí, které jsou přístupné pomocí IIIF. Pomocí tohoto aplikačního rozhraní je možné se doptat kompatibilního repositáře, jaké kolekce nabízí, které dokumenty kolekce obsahují a teoreticky je možné takto přejímat celé komplexní digitální dokumenty v nezkrácené podobě. Dosavadní praxe IIIF komunity je spíše taková, že předávána jsou elementární popisná metadata a co nejucelnější metadata strukturální (fyzické a logické mapy).

V okamžiku, kdy jsme vyřešili problematiku přístupu k jednotlivým distribuovaným obrazům prostřednictvím jednotného IIIF kompatibilního rozhraní Manuscriptoria, je dalším logickým krokem právě implementace IIIF Presentation API.

3.3.1 IIIF Presentation API při poskytování přístupu k agregovanému obsahu

Výroba tzv. manifestů tedy byla zařazena do procesů zpracování agregovaných metadat a manifesty dokumentů⁵ jsou zpřístupňovány tak, aby s obsahem Manuscriptoria mohly nakládat zainteresované projekty třetích stran, nebo dokonce i jednotliví uživatelé, a to i mimo uživatelské prostředí Manuscriptoria s pomocí libovolných alternativních zobrazovacích nástrojů.

Jednotlivé kolekce postupně zveřejňujeme již v plném provozu v rámci poskytování služeb Manuscriptoria v. 3.0 prostřednictvím IIIF Universe

(<https://graph.global/static/data/universes/iiif/universe.json>).

⁵ Manifest obsahuje informace o obsahu jednoho dokumentu v předepsané formě.



Obrázek 4: Manuscriptorium v IIIF Universe (na pozici 13)



```
JSON  Surová data  Hlavičky
Uložit  Kopírovat
@context: "http://iiif.io/api/presentation/2/context.json"
@id: "http://collectiones.manuscriptorium.com/"
@type: "sc:Collection"
label: "Manuscriptorium"
viewingHint: "top"
description: "Highlighted Collections"
attribution: "Collection provided by Manuscriptorium"
collections:
  0:
    @id: "http://collectiones.manuscriptorium.com/unescoCzechReformation/"
    @type: "sc:Collection"
    label: "Collection of medieval manuscripts of the Czech Reformation"
    license: "https://creativecommons.org/licenses/by-nc-nd/4.0/"
  1:
    @id: "http://collectiones.manuscriptorium.com/unescoUniversityTheses/"
    @type: "sc:Collection"
    label: "Collection of 526 prints of university theses from 1637-1754"
    license: "https://creativecommons.org/licenses/by-nc-nd/4.0/"
```

Obrázek 5: První kolekce Manuscriptoria v IIIF Universe

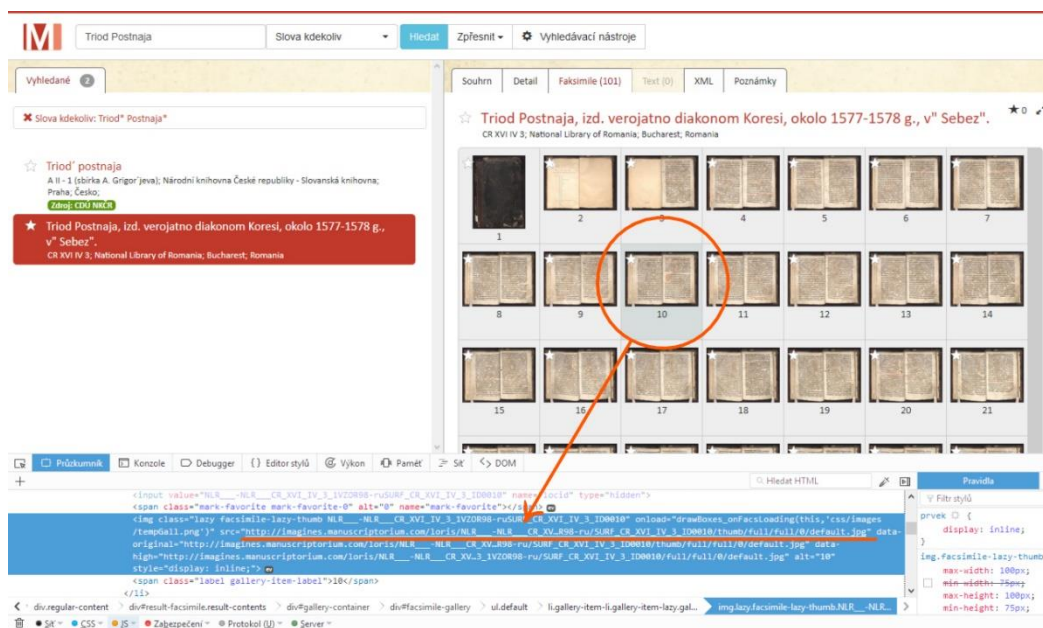
Dalším krokem vývojových prací pak bude přidání manifestů dokumentů do existujících OAI-PMH profilů Manuscriptoria tak, aby s obsahem mohl nakládat například projekt EUROPEANA a další partneři.

3.3.2 Využití IIIF Presentation API při přístupu k strukturálním metadatům distribuovaných repositářů kompatibilních s IIIF

V režimu poloprovozu je testována funkcionalita načítání IIIF manifestů z kompatibilních repositářů. Tato funkcionalita, kdy partnerský repositář poskytuje odkazy na IIIF manifesty namísto proprietárních strukturálních metadat, bude brzy uvedena plného provozu (ověřovací implementace probíhá s využitím IIIF kolekcí projektu e-Codices (<http://www.e-codices.unifr.ch/en>)).

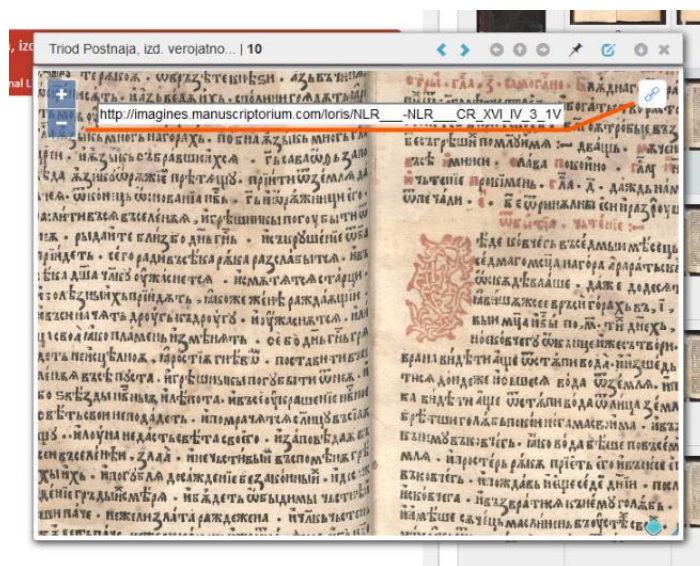
4 Ukázky využití IIIF v prostředí Manuscriptoria

Manuscriptorium při zpřístupňování obrazů plně přešlo na komunikaci prostřednictvím IIIF Image API. Tato změna je zásadní, ale nezasvěcenému uživateli je na první pohled skrytá. Teprve při hlubším zkoumání se můžeme přesvědčit, že tomu tak skutečně je.




Obrázek 6: Pohled pod pokličku UI Manuscriptoria ve vývojářské konzoli

Běžný uživatel se s IIIF odkazováním setká, pokud si zobrazí permanentní link na zobrazovaný obrázek.











Obrázek 7: Permanentní link na obrázek dle IIIF




Dále může uživatel pracovat s nástrojem pro manipulaci s obrazem a výsledek sdílet pomocí odkazu dle IIIF – takto může zahrnout obraz do vlastního blogu, seminární práce apod. Nástroj se otevírá pomocí ikony editace  v horní liště aktivního obrázku.

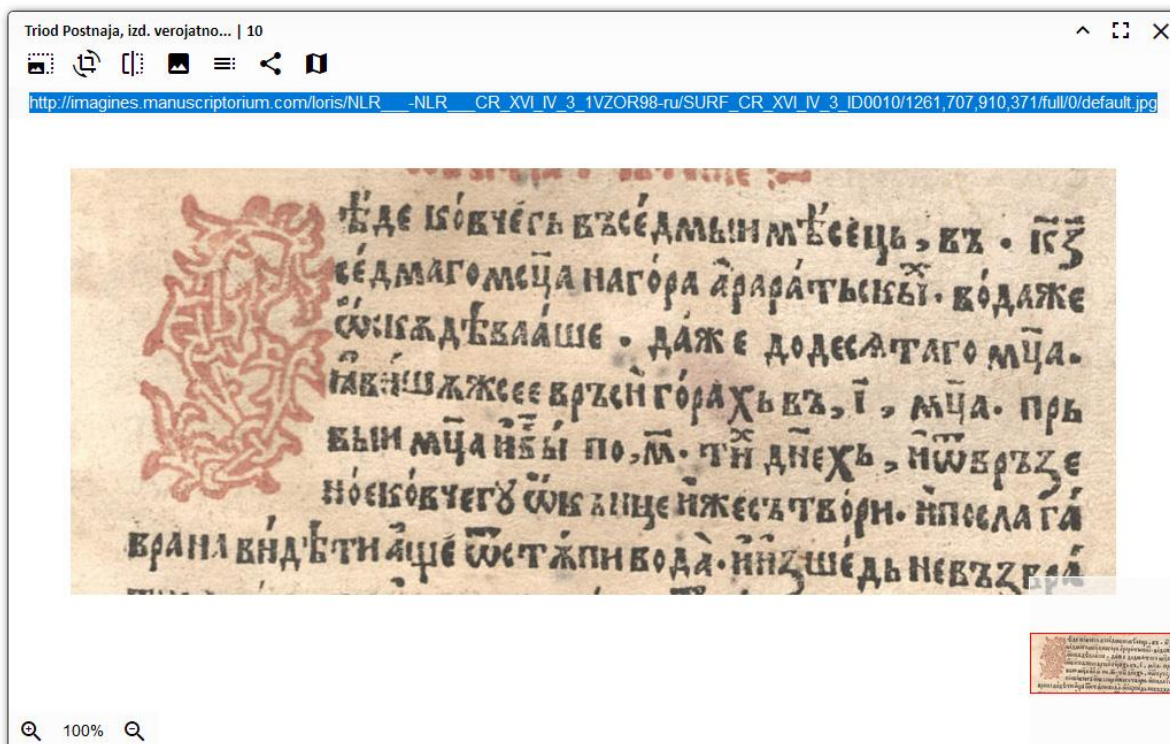
Editační prostředí umožňuje

- Vybírat kvalitu zobrazení (z předdefinovaných kvalit)  ,
- měnit velikost (převzorkování) obrazu  nebo  100%  ,
- vybrat a oříznout oblast obrazu  ,
- otáčet obraz  ,
- překlopit obraz  .

Veškeré modifikace se v prostředí potvrzují kliknutím na ikonku Potvrzení  .

Editor je také patřen historií úprav, kterou lze vyvolat kliknutím na tlačítko Historie  , lze se tedy v úpravách vracet po jednotlivých krocích.

Výsledek manipulace může být sdílen pomocí IIIF URL, které lze kopírovat po stisknutí tlačítka Sdílení



Obrázek 8: ukázka práce s obrazem



5 Závěr

Vývojovému týmu se podařilo do praxe uvést jedinečné řešení, které v oblasti agregace historických dokumentů nemá dosud obdoby. Hladký průběh řešení a analýz byl umožněn také díky podpoře ze strany komunity IIF – rozsah spolupráce členů a kvalita dostupných nástrojů, zdrojů a rozsah sdílených znalostí dostupných on-line dává záruku, že zvolený standard a směr budoucího vývoje jsou správné a přiměřeně životné.