

IMPLEMENTACE IIIF PRESENTATION API BRÁNY PRO SYSTÉM KRAMERIUS

Ing. Petr Žabička; petr.zabicka@mzk.cz; (Moravská zemská knihovna v Brně)

Mgr. Pavla Rychtářová; pavla.rychtarova@mzk.cz; (Moravská zemská knihovna v Brně)

Rodina mezinárodních standardů pro zpřístupnění digitalizovaných dokumentů IIIF se od svého vzniku před deseti lety vyvinula ve významnou aktivitu, kolem které existuje živá otevřená komunita významných hráčů na poli zpřístupnění digitálních dokumentů. Systém Kramerius je již řadu let vybaven rozhraním (API), které poskytuje všechny relevantní informace, vyžadované standardy IIIF. Tento příspěvek popisuje aplikaci, která funguje jako brána mezi rozhraním Krameria a standardem IIIF Presentation API a otevírá tak digitální knihovny postavené na Krameriu světu IIIF.

<http://doi.org/10.52036/1335793X.2023.1-2.9-13>

Jednou z hlavních rolí paměťových institucí je vedle uchování informací zejména jejich šíření. To bylo výrazně usnadněno nástupem internetu a digitalizací, v mnoha případech je však přístup uživatelů k informacím komplikován i jejich roztržitostí. Uživatel je postaven před velké množství různých systémů, z nichž každý má specifické uživatelské rozhraní a technické možnosti pro další práci se zpřístupněnými dokumenty. Ačkoli by se mohlo zdát, že heterogenita systémů pro zpřístupnění digitalizovaných dokumentů je prospěšná – motivuje tvůrce těchto systémů ke vzájemnému soutěžení při vylepšování svých aplikací –, z hlediska uživatele je opak pravdou. Ačkoli se může uživatel naučit pracovat s řadou různých systémů, zvláště pokud je logika práce s nimi obdobná, dříve nebo později narazí na různé limity a omezení jednotlivých systémů. To se projevuje zejména v případě, kdy je cílem další využití digitalizovaných dokumentů – například v podobě online výstav, vytváření virtuálních sbírek nebo jiném odborném zpracování digitalizátů.

IIIF IMAGE API

Výše naznačený problém vedl v roce 2011 několik nejvýznamnějších světových knihoven a univerzit k vytvoření sady standardů pro sdílení informací, nazvaném IIIF (International Image Interoperability Framework). Jak již název napovídá, primárním cílem bylo umožnit přímé sdílení obrazových dat prostřednictvím k tomu vytvořeného standardu – IIIF Image API (Image nedatováno). Tento standard velmi usnadňuje práci s jednotlivými online zpřístupněnými obrázky. Prostřednictvím jednotně definované syntaxe URL je tak možné zpřístupnit digitalizovaný dokument (např. obraz nebo stránku z knihy) v podobě požadované uživatelem. Ten

může jen úpravou odkazu získat daný digitalizát nebo jeho výřez v požadovaném rozlišení, barevnosti, otočení nebo datovém formátu.

IIIF Image API tak během uplynulého desetiletí zahájilo úspěšné tažení světem, jeho podpora se stává v digitálních knihovnách samozřejmostí. V České republice je prostřednictvím IIIF Image API zpřístupněna například celá Národní digitální knihovna v systému Kramerius, IIIF Image API podporuje i systém Manuscriptorium.

IIIF PRESENTATION API

Snadným sdílením jednotlivých snímků ale nelze považovat otevření systémů digitálních knihoven za dokončené. Zejména tam, kde je součástí zpřístupněného díla více než jeden snímek, je nutné předat i informaci o tom, jaký je vztah mezi jednotlivými snímky coby částmi daného díla. K tomu je nezbytné poskytnout řadu údajů – metadata popisující jak celé dílo, tak jednotlivé části díla a jejich vzájemný vztah (např. pořadí stránek v knize) a případně další informace, jako například anotace. IIIF pro tento účel definuje další standard – Presentation API. To je navrženo tak, aby jednotným způsobem poskytlo všechny klíčové informace nezbytné pro další práci s digitalizovaným dokumentem.

IIIF Presentation API tak umožňuje online aplikacím prezentovat uživatelům i komplexní digitální objekty obohacené o řadu souvisejících informací a vazeb. Tím, že toto API předává informace určené lidem, jsou i popisné informace uvedeny způsobem vhodným pro čtení lidmi. Specifikace tohoto API explicitně uvádí, že není zaměřeno na poskytování metadat, která by umožnila vyhledávací indexovat digitální objekty

– k tomu slouží jiné standardy, jako například OAI-PMH, nebo ze standardu ResourceSync vycházející IIIF Change Discovery API.

Využití tohoto API tak umožní aplikacím zobrazit uživateli digitalizované obrázky, video, audio a další typy obsahu spojené s konkrétním fyzickým nebo digitálním objektem a procházet mezi nimi ať už postupně, nebo hierarchicky. Zároveň umožní nabídnout uživateli popisné, strukturální nebo další kontextové informace, odkazy na další zdroje a případně i aktivní interakci uživatele s objektem. Takovou interakcí může být například sdílení tvoření anotací, přepis textu nebo zvuku, přiřazování geografických souřadnic apod. Je nutné uvést, že toto API neposkytuje interaktivní služby, jako je například prohledávání dokumentu (tuto problematiku řeší IIIF Content Search API), nebo autentizaci (k tomu je určeno IIIF Authentication API).

IIIF Presentation API je velmi podrobně popsáno přímo ve své specifikaci (Presentation nedatováno). Je postaveno na souboru metadat ve formátu JSON, který je odpovědí na požadavek ve formě URI. API má definováno několik typů metadatových souborů, z nichž tím základním je tzv. manifest, typ objektu popisující základní intelektuální entitu, například knihu nebo číslo periodika. Dalším důležitým typem objektu je sbírka (collection), která může obsahovat jak manifesty, tak další sbírky. Například titul periodika je konstruován jako sbírka ročníků, každý ročník je pak sbírkou manifestů jednotlivých čísel. Běžný manifest pak obsahuje sadu pláten (canvas), na která se umísťují jednotlivé stránky.

Díky relativně statickému charakteru Presentation API je možné k jeho implementaci přistupovat dvojím způsobem: buď je možné vytvořit dávkovou proceduru, která bude pravidelně spouštěna nad digitální knihovnou a bude generovat soubory IIIF sbírek a manifestů, nebo naprogramovat dynamickou aplikaci, která bude tyto soubory generovat v reálném čase na základě obsahu digitální knihovny nebo jejího indexu.

KRAMERIUS A IIIF

Systém Kramerius je s několika desítkami instalací nejrozšířenější aplikací digitální knihovny v České republice, používanou téměř všemi knihovnami, které provádějí digitalizaci svých fondů. Ze zahraničních instalací Krameria je nutné zmínit Slovenskou národní knihovnu, která je zároveň jednou z největších instalací, co se množství zpřístupňovaných titulů týče. Jedním z důsledků velkého množství instalací Krame-

ria je i velká roztržitost, co se jeho verzí týče. Zatímco některé knihovny již přešly nebo přecházejí na nejnovější verzi 7, řada knihoven provozuje i šest nebo více let staré verze Krameria 5 (zcela pomíjíme několik stále přežívajících instalací Krameria 3, kterým už táhne na dvacátý rok). Tím, co odlišuje Kramerius 5 od jeho starších verzí, je existence API, které poskytuje všechny funkce potřebné pro zpřístupnění obsahu Krameria veřejnosti. Kramerius 5 se tak stal první verzí, která umožnila plně oddělit serverovou část Krameria od jeho webové podoby, přístupné uživatelům. Toto API, které fakticky plní stejnou roli jako rodina standardů IIIF, umožnilo obrovský skok v uživatelské přívětivosti Krameria a výrazně usnadnilo jeho rozvoj, protože webový klient Krameria dokáže fungovat jak nad nejnovějšími verzemi Krameria 7, tak i nad nejstaršími instalovanými verzemi Krameria 5, což dokládá mimo jiné i webový portál Digitalniknihovna.cz, který umožňuje uživatelům pracovat s významnou většinou instalací Krameria z jednoho místa.

Vývoj API Krameria započal přibližně ve stejné době, kdy bylo založeno konsorcium IIIF, a probíhal nezávisle na něm. Zatímco cílem API Krameria bylo poskytnout plnou funkčnost digitální knihovny se specifickým zaměřením na digitalizaci v českých knihovnách, rodina IIIF API byla zaměřena na definování minimální sady informací pro zajištění interoperability mezi různými poskytovateli dat na straně jedné a prohlížečkami dat na straně druhé.

Implementací podpory IIIF Presentation API (tehdy platné verze 2) v Krameriu se MZK začala zabývat v roce 2015. Využila toho, že imageserver používaný v Krameriu (IIPsrv) nabízel i IIIF Image API a vyvinula rozšíření Krameria, schopné poskytovat IIIF manifesty odpovídající specifikaci Presentation API. Tyto manifesty však bylo možné generovat jen pro dokumenty obsahující stránky, chyběla tak podpora pro zvukové dokumenty, vícesvazková díla nebo tituly a ročníky periodik. Již v této první verzi implementace bylo rozhodnuto, že IIIF manifesty musejí být generovány dynamicky na základě požadavku uživatele. I když se tím do jisté míry zvyšuje zátěž serveru, eliminuje se tím nutnost udržovat na serveru aktuální kopii všech manifestů, což zvyšuje rizika toho, že požadovaný manifest nebude aktuální, případně nebude vůbec vygenerovaný zrovna v okamžiku, kdy někdo požádá o jeho stažení.

S postupným dalším vývojem Krameria se navíc začaly ukazovat i další nevýhody tohoto řešení: protože se

jednalo o součást jádra Krameria, byla instalace tohoto rozšíření závislá na instalaci celé nové verze Krameria v jednotlivých knihovnách. Opravy a drobná vylepšení, nebo i změny vyvolané aktualizacemi specifikace Presentation API (verze 3.0 byla publikována v polovině roku 2020) bylo sice možné nasadit v MZK, ale bylo téměř nemožné je šířit do dalších knihoven. Dalším impulsem vedoucím k rozhodnutí o nové implementaci API pak byl připravovaný přechod na Kramerius verze 7, který se od verze 5 liší nejen tím, jak ukládá svá data interně, došlo i ke změnám ve struktuře indexu a veřejného API. Kramerius 7 navíc přinesl nové řešení virtuálních sbírek, které se nově chovají jako specifický druh dokumentu, který může obsahovat jak jiné dokumenty, tak další sbírky a velmi se tak blíží tomu, jak se na sbírky dívá IIIF.

IMPLEMENTACE IIIF BRÁNY

MZK stála tedy před rozhodnutím, jak zajistit podporu IIIF Presentation API v Krameriiích. Výše zmíněné negativní zkušenosti s implementací IIIF API jako součástí jádra Krameria vedly ke dvěma možným variantám: poskytovat toto rozšíření knihovnám podobně jako webový klient Krameria, tedy jako aplikaci, kterou si každá knihovna nainstaluje na svůj server a bude ji provozovat spolu s Krameriem, nebo jít cestou obdobnou citační službě pro Krameria, která funguje centrálně pro všechny Kramerie, je ji možné snadno aktualizovat nebo opravit a i díky tomu funguje velmi spolehlivě. Obdobně pozitivní zkušenosti měla již MZK s provozováním portálu Digitalniknihovna.cz, který slouží nejen jako místo, odkud je možné přistupovat do většiny instalovaných Krameriiů, ale pro Moravskou zemskou knihovnu v Brně a Městskou knihovnu v Praze slouží jako hlavní přístupový bod do jejich digitálních knihoven.

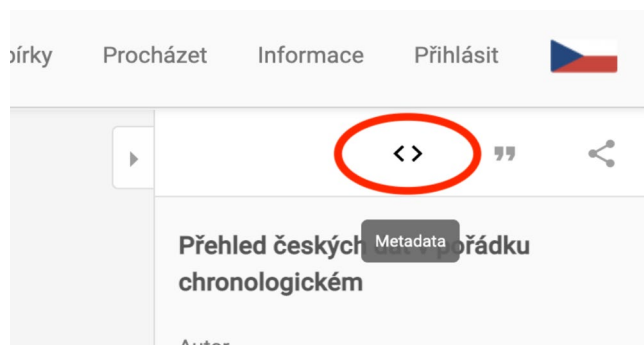
Vzhledem k tomu, že jedním z hlavních cílů vývoje IIIF rozhraní byl záměr poskytnout je co nejširšímu počtu

Krameriiů, byla zvolena naposledy uvedená varianta, tedy centrální systém, fungující jako brána mezi API jednotlivých Krameriiů na straně jedné a IIIF Presentation API na straně druhé. Systém je postaven jako webová aplikace, jejíž backend je vytvořen v prostředí Node.js v kombinaci s frameworkem Express, s frontendem běžícím na adrese iiif.digitalniknihovna.cz. Jádrem aplikace je Ruby skript, který na základě požadavku vytvoří z metadat získaných prostřednictvím API Krameria výstup ve formátu IIIF Presentation API. Aplikace podporuje jen systémy Kramerius verzí 5 a 7, které musí běžet na https. Po přechodu na Kramerius 7 tak bude využití IIIF brány otevřeno i pro online digitální knihovnu Slovenské národní knihovny.

Implementace IIIF Presentation API v Krameriu vychází z metodiky (Žabička et al. 2020), kterou MZK připravila v roce 2021. Aplikace má dva základní režimy: V tom prvním, standardním, je aplikaci předán identifikátor Krameria (podle Registru Krameriiů) a identifikátor objektu z tohoto Krameria. Na základě těchto identifikátorů pak aplikace zašle na API daného Krameria sérii dotazů a vrátí vygenerovaný IIIF manifest. Jako experimentální rozšíření je k dispozici možnost aplikaci předat jen link do daného Krameria a aplikace sama z tohoto linku vyextrahuje potřebné identifikátory a vygeneruje manifest. Tato druhá varianta je v současnosti podporovaná jen pro odkazy z portálu Digitalniknihovna.cz.

Zobrazit IIIF manifest je možné přímo z webového rozhraní Krameria prostřednictvím tlačítka Metadata (viz obr. 1) a následně volbou IIIF z první položky menu (viz. obr. 2). URL zobrazené v pravém horním rohu okna s manifestem je pak možné zkopírovat a použít v jiné aplikaci podporující IIIF Presentation API.

IIIF brána zobecňuje strukturální specifika různých modelů dokumentů v Krameriu a vytváří na jejich zá-



Obr. 1 Tlačítko Metadata



Obr. 2 Zobrazení IIIF manifestu

kladě buď IIIF manifesty, nebo IIIF kolekce. Mimo již zmíněných virtuálních sbírek z Krameria 7 jsou IIIF kolekce tvořeny z titulů a ročníků periodik, z vícesvazkových děl monografického charakteru a z konvolutů, ostatní modely (včetně zvukových dokumentů) převádí na IIIF manifesty. Pokud pro příslušný dokument nejsou k dispozici obrázky stránek prostřednictvím IIIF Image API, jsou do manifestu generovány odkazy na statické snímky stránek tak, jak je zobrazuje daný Kramerius (obvykle se jedná o jpg). Specifické je řešení dokumentů typu pdf, kdy se v IIIF manifestu předávají jen metadata obsahující odkaz na samotné pdf.

Zobecnění jednotlivých modelů Krameria, samozřejmě, vede k určité ztrátě informace a smývá se tak například rozdíl mezi vícesvazkovým dílem a konvolutem nebo virtuální sbírkou a periodikem. Základní struktura dokumentu je však i v takovém případě zachována a pro uživatele zůstává s využitím popisných metadat dostatečně srozumitelná.

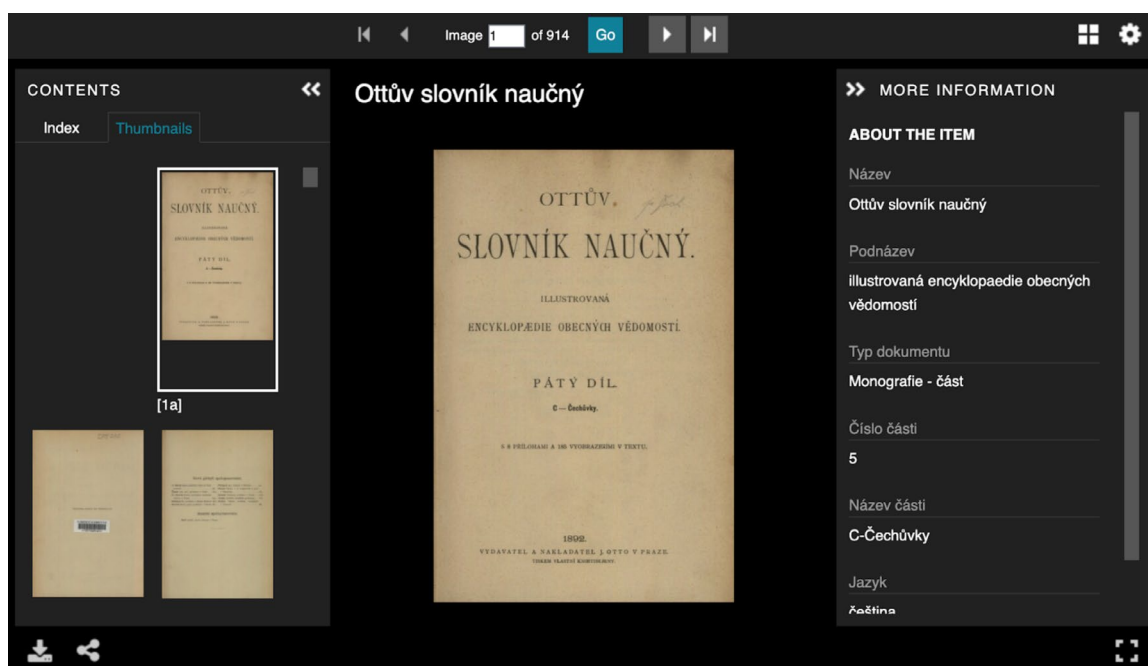
Největší implementační výzvu však nepředstavovaly kolekce, ale rozsáhlé, více než tisícistránkové svazky, například některé středověké rukopisy nebo rozsáhlé tištěné dokumenty z 19. nebo první poloviny 20. století. Jedním z povinných údajů při tvorbě IIIF manifestu jsou totiž rozměry jednotlivých stránek, které se používají k definici velikosti virtuálního „plátna“, které uvidí uživatel a na které se má obrázek stránky nanášet.

Velikost obrázku však není v Krameriu nijak zaindexována a je možné jej získat jedině dotazem na vlastnosti obrázku prostřednictvím IIIF Image API. 1000x opakovaný dotaz na rozměry každé stránky dokumentu však může velmi zpomalit odezvu serveru, dokonce natolik, že před sestavením manifestu dojde k timeoutu a uživatel manifest vůbec nezíská. Ani Kramerius 7 tento problém z technických důvodů neřeší, protože indexace a reindexace dokumentů v něm by trvala podstatně déle než nyní, což je při správě velkých digitálních knihoven nepřijatelné. IIIF brána proto tento problém řeší tak, že se na velikost obrázků dotazuje pro prvních 50 stránek dokumentu a pro další stránky generuje velikosti plátna shodně bez dalších dotazů do Krameria.

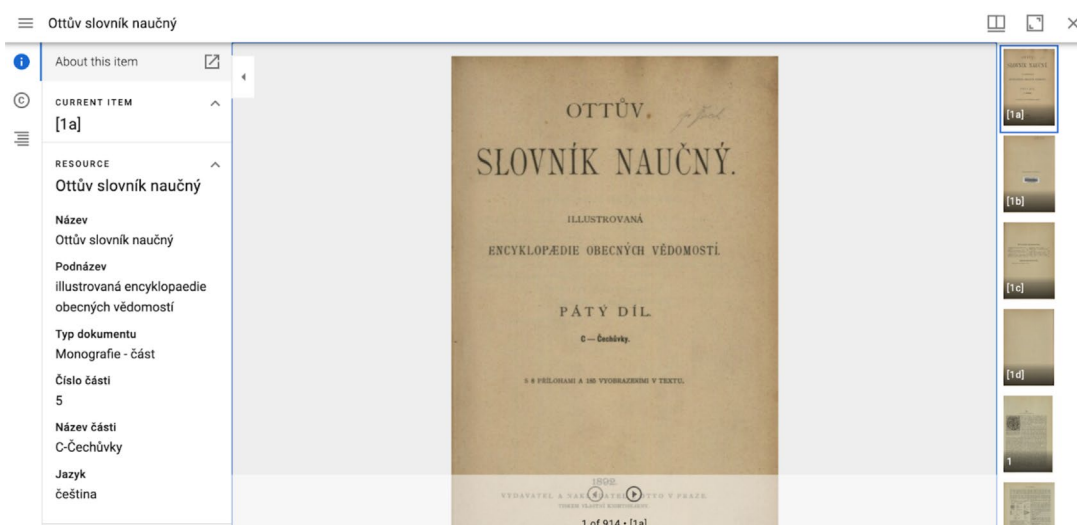
VYUŽITÍ IIIF PRESENTATION API

Pro přímé zobrazení dokumentů prostřednictvím IIIF Presentation API je možné použít mimo jiné volně dostupné prohlížečky Mirador (nedatováno) a Universal Viewer (nedatováno). Obě podporují jak manifesty, tak kolekce, obě však mají určité problémy například s prezentací či přehráváním zvukových dokumentů z Krameria.

Zatímco Universal Viewer (obr. 3) je vhodný pro čtení jednoho dokumentu, Mirador (obr. 4) prohlížečka je optimalizovaná pro komparativní práci, při které uživatel může paralelně listovat a porovnávat několik dokumentů zároveň.



Obr. 3 Prohlížečka Universal Viewer



Obr. 4 Prohlížečka Mirador

Prohlížečka Universal Viewer je tak vhodnější například pro vložení dokumentu do webové stránky a díky tomu k prezentaci dokumentu z digitální knihovny v jiném kontextu (např. virtuální výstava). Díky tomu, že tuto prohlížečku používá pro prezentaci obsahu své digitální knihovny Wellcome Library, je tato prohlížečka poměrně propracovaná i co se uživatelské přívětivosti týče (Henshaw 2011).

Mirador (obr. 4) je naproti tomu jednoznačně nástrojem pro pokročilého badatele (Van Zundert 2018). Je možné jej doplnit ovládacími prvky pro modifikace parametrů grafického zobrazení obrázku, sdílené anotace a mnoha dalšími. Je otevřený i pro přidávání dalších pluginů, které si uživatelé mohou sami naprogramovat a přizpůsobit si jej tak vlastní odborné práci, při které mohou vedle sebe postavit dokumenty z řady různých digitálních knihoven z celého světa a nemusí řešit specifika jejich uživatelského rozhraní, technických omezení apod.

Zobrazení v jiných prohlížečkách, než jsou ty poskytované samotnou digitální knihovnou, však není jediná možnost využití IIIF Presentation API. Přestože je toto rozhraní primárně určeno lidským uživatelům, jeho unifikovanost otevírá nové možnosti například i v oblasti strojového učení (Hankiewicz 2020), kde bude díky standardizaci snadnější vytvářet například specializované datové sady ať už pro tvorbu nových modelů, nebo pro strojové obohacování dokumentů obsažených v digitálních knihovnách.

Studie vznikla v rámci Dlouhodobé koncepce pro rozvoj výzkumné organizace – Moravská zemská knihovna v Brně.

POUŽITÁ LITERATURA

- HANKIEWICZ, Steffen a Oliver PAETZEL, 2020. Machine Learning and IIIF in the Reality Check of Daily Digitization Projects using the Example of the Goobi Community. *Archiving Conference* [online]. 17(1), 72–78 [cit. 2022-12-08]. ISSN 2161-8798. Dostupné z: doi: [10.2352/issn.2168-3204.2020.1.0.72](https://doi.org/10.2352/issn.2168-3204.2020.1.0.72)
- HENSHAW, Christy, 2011. The Wellcome Digital Library: building a sustainable infrastructure. *Library Hi Tech News* [online]. 28(1), 18–21 [cit. 2022-12-08]. ISSN 0741-9058. Dostupné z: doi: [10.1108/074190511111130394](https://doi.org/10.1108/074190511111130394)
- Image API 3.0* [online]. The IIIF Consortium, 2022 [cit. 2022-12-10]. Dostupné z: <https://iiif.io/api/image/3.0/>
- Mirador* [online]. [b.r.]. Dostupné z: <https://projectmirador.org/PresentationAPI3.0> [online]. The IIIF Consortium, 2022 [cit. 2022-12-10]. Dostupné z: <https://iiif.io/api/presentation/3.0/>
- Registr Kramériů* [online]. [cit. 2022-12-10]. Dostupné z: <https://registr.digitalniknihovna.cz/>
- UniversalViewer*. In: GitHub [online]. The Universal Viewer Steering Group, [b.r.]. Dostupné z: <https://github.com/UniversalViewer/universalviewer/wiki/About>
- VAN ZUNDEERT, Joris, 2018. On Not Writing a Review about Mirador: Mirador, IIIF, and the Epistemological Gains of Distributed Digital Scholarly Resources. *Digital Medievalist* [online]. 11(1) [cit. 2022-12-10]. ISSN 1715-0736. Dostupné z: doi: [10.16995/dm.78](https://doi.org/10.16995/dm.78)
- ŽABIČKA et al., 2020. *Metodika implementace IIIF v českých digitálních knihovnách: se zvláštním zřetelem na systém digitální knihovny Kramerius*. Brno: Moravská zemská knihovna. Dostupné také z: <http://www.nusl.cz/ntk/nusl-454347>

■ Článek bol recenzovaný