

OAI Adapter - uživatelský manuál

[Úvod](#)

[Uživatelské účty](#)

[Vstupní parametry](#)

[Kód registrátora](#)

[Mód registrace](#)

[Základní OAI URL](#)

[OAI metadatový prefix](#)

[OAI set](#)

[Transformační šablony](#)

[Úpravy a testování šablon](#)

[Na co si dát pozor](#)

[Report.txt a logy](#)

[Spuštění přes webové rozhraní](#)

[Spouštění z příkazové řádky](#)

[Configuration.properties](#)

[O tomto dokumentu](#)

Úvod

OAI Adapter je nástroj, který slouží k usnadnění hromadné a pravidelné registrace digitálních dokumentů (přiřazení URN:NBN, dále jen *registrace DD*) a importu digitálních instancí (přiřazení URL k URN:NBN, dále jen *import DI*) v systému CZIDLO.

Pracuje tak, že sklízí metadatové záznamy pomocí protokolu OAI-PMH 2.0 z OAI Repozitářů, typicky samotných digitálních knihoven. Záznamy transformuje a následně používá jako vstupy operací registrace DD a import DI přes CZIDLO API.

OAI Adapter je možné spouštět dvojím způsobem. Buď z terminálu/příkazového řádku, nebo jak proces na serveru spravovaný přes webové rozhraní.

Více informací naleznete zde:

<https://github.com/NLCR/CZIDLO/wiki/OAI-Adapter>

Uživatelské účty

Pro spuštění OAI Adapteru je nutné použít platný uživatelský účet. Uživatel také musí mít přiřazena práva ke konkrétnímu registrátorovi. Za tvorbu účtů a přiřazování práv je odpovědný administrátor nebo kurátor. Kontakt na něj je k dispozici na webu <http://resolver.nkp.cz> v záložce **Informace**.

Vstupní parametry

Vstupní parametry pro běh OAI Adapteru jsou definovány buď v souboru `.properties`, nebo v případě spuštění z webového rozhraní vyplňovány do formuláře dialogu. Parametry jsou následující:

OAI Adapter - uživatelský manuál

Kód registrátora

Kód registrátora, jemuž digitální dokumenty patří. Kód je součástí samotného URN:NBN. Při práci s více registrátory je potřeba spouštět OAI Adapter pro každého z nich zvlášť.

Mód registrace

Definuje, zda mají být identifikátor URN:NBN přiřazován na serveru systémem (BY_RESOLVER), nebo spravován registrátorem samotným a systémem pak jen potvrzován (BY_REGISTRAR), případně vybírán registrátorem, ale ze seznamu dříve alokovaných identifikátorů (BY_RESERVATION).

Ve většině případů bude mít registrátor povolen jeden mód registrace. Nicméně technicky je možné módy registrace kombinovat. Pokud toto nastane a zároveň jsou metadata různých dokumentů zpracována v jednom běhu OAI Adapteru, může vypadat report poněkud zmateně. Např. pokud při spouštění s mód_registrace=BY_REGISTRAR narazí na dokument, který již URN:NBN přiřazené má a to je ve stavu VOLNÉ, je zřejmé, že se buď jedná o chybu v metadatach, anebo by tento dokument měl být zpracován v módu BY_REGISTRAR, nebo BY_RESERVATION. Každopádně je záznam vyhodnocen jako chybový. Proto je nutné v takovém případě spouštět OAI Adapter opakovaně a přepínat mezi obsahem parametru *mód registrace*.

Základní OAI URL

Základní URL OAI Poskytovatele¹. Např. <http://kramerusdemo.mzk.cz/oaiprovider>. Pro kontrolu stačí za základní URL připojit řetězec “?verb=Identify”, tedy <http://kramerusdemo.mzk.cz/oaiprovider?verb=Identify>, a otevřít ve webovém prohlížeči. Výsledkem by měl být zobrazený popis repozitáře v xml.

OAI metadatový prefix

Identifikátor formátu metadatové reprezentace sklizených záznamů. Např. *oai_dc* (Dublin Core²), *mods* (MODS), *oai_marc* (MARC21), *drkramerus4* (interní formát Krameria), *czidlo* (interní formát CZIDLO), *ese* (formát Europeany). Seznam podporovaných formátů lze získat následovně: <http://kramerusdemo.mzk.cz/oaiprovider?verb=ListMetadataFormats>. Transformační šablony jsou vyrobeny na míru metadatového formátu, resp. jeho konkrétnímu dialektu používanému v repozitáři.

OAI set

Set (množina), tedy omezený soubor záznamů z OAI. Např. Kramerus používá množiny *periodical*, *monograph*, *collection*, apod. Množina je nepovinný parametr, prakticky ale většinou vyplněn bývá. Množiny jsou totiž častým nástrojem, jak oddělit jednotlivé typy dokumentů (např. monografie a periodika v Krameriu). Ze samotných metadat už pak nemusí být zřejmé, o jaký typ dokumentu jde. Proto dává smysl množiny vyplňovat a používat specifické transformační

¹ OAI Poskytovatel (Provider) - API pro sklizení metadat z OAI Repozitáře.

² Dublin core je povinný, každý OAI Poskytovatel musí vracet záznamy v tomto formátu.

OAI Adapter - uživatelský manuál

šablony pro každou množinu. A OAI Adapter pak spouštět pro každou množinu zvlášť. Seznam množin se dá zjistit takto: <http://kramerusdemo.mzk.cz/oaiprovider?verb=ListSets>.

Transformační šablony

Transformační šablony XSLT převádějící metadatový záznam sklizený od OAI Poskytovatele na dvě xml, která jsou vstupem operací *registrace DD* a *import DI*. Obě šablony na vstupu očekávají záznam v konkrétním metadatovém formátu a na výstupu vrací záznam ve formátu CZIDLO. Některá data mohou být přímo součástí šablony³. Typicky kořenový element (monograph/periodical/periodicallssue, ...) pro vstup *registrace DD*, nebo id digitální knihovny pro *import DI*.

Příklady šablon:

- monografie ve formátu Kramerius:
https://github.com/NLCR/CZIDLO/tree/master/examples/oaiAdapter_xslt/kramerus/monograph
- mapy ve formátu MARC21:
https://github.com/NLCR/CZIDLO/tree/master/examples/oaiAdapter_xslt/oldMaps
- články časopisu Duha ve formátu Dublin Core:
https://github.com/NLCR/CZIDLO/tree/master/examples/oaiAdapter_xslt/duha
- všechny příklady:
https://github.com/NLCR/CZIDLO/tree/master/examples/oaiAdapter_xslt

Vždy jsou zde soubory **NAZEV_OAI-FORMAT.xml** (původní záznam z OAI), **NAZEV_OAI-FORMAT_to_czidloDD.xsl** a **NAZEV_OAI-FORMAT_to_czidloDI.xsl** (samotné šablony) a **NAZEV_czidloDd.xml** a **NAZEV_czidloDi.xml** (pomocí šablon transformované záznamy, tedy vstupy operací *registrace DD* a *import DI*).

Úpravy a testování šablon

Pro tvorbu a úpravy transformačních XSLT šablon budete potřebovat editor XML, který podporuje XSLT 1.0 a ideálně i validaci s pomocí XSD schémat. Porovnání možností některých editorů: https://en.wikipedia.org/wiki/Comparison_of_XML_editors.

XSD je potřeba z toho důvodu, že vstupy operací se na serveru validují podle XSD dostupných zde: <https://github.com/NLCR/CZIDLO/tree/master/api/src/main/resources/v3>. Proto je vhodné i během úprav šablon průběžně validovat výstupy.

Pro základní transformace a validace mohou postačit online nástroje, třeba:

<http://www.freeformatter.com/xsl-transformer.html>

<http://www.freeformatter.com/xml-validator-xsd.html>

Na co si dát pozor

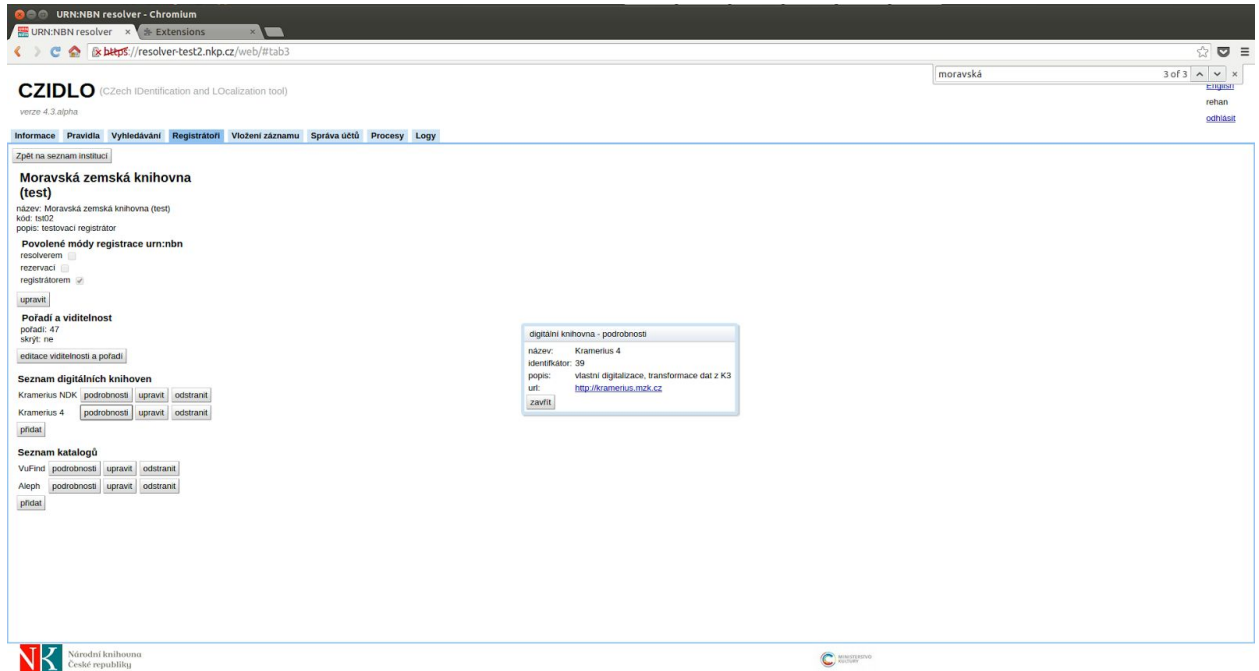
Při tvorbě úpravách šablon je potřeba si dávat pozor zejména na následující elementy ve výstpních xml:

- *digitalLibraryId (import DI)* - Číselný identifikátor digitální knihovny, v níž se digitální instance nacházejí. CZIDLO eviduje digitální knihovny každého registrátora. Těch může být více (Kramerius 3, Kramerius 4, vlastní webová prezentace, ...). Typicky budou

³ Více v kapitole *Na co si dát pozor*.

OAI Adapter - uživatelský manuál

všechny záznamy zpracované v jednom běhu OAI Adapteru ve stejné digitální knihovně. Proto bývá *digitalLibraryId* v šabloně na pevně. Identifikátor digitální knihovny zjistíme přes webové rozhraní v záložce **Registrátoři** po rozkliknutí **Podrobnosti**. V sekci **Seznam digitálních knihoven** můžeme digitální knihovnu vytvořit tlačítkem **přidat**. Kliknutím na tlačítko **podrobnosti** u digitální knihovny se otevře dialog, který mimojiné obsahuje identifikátor digitální knihovny.



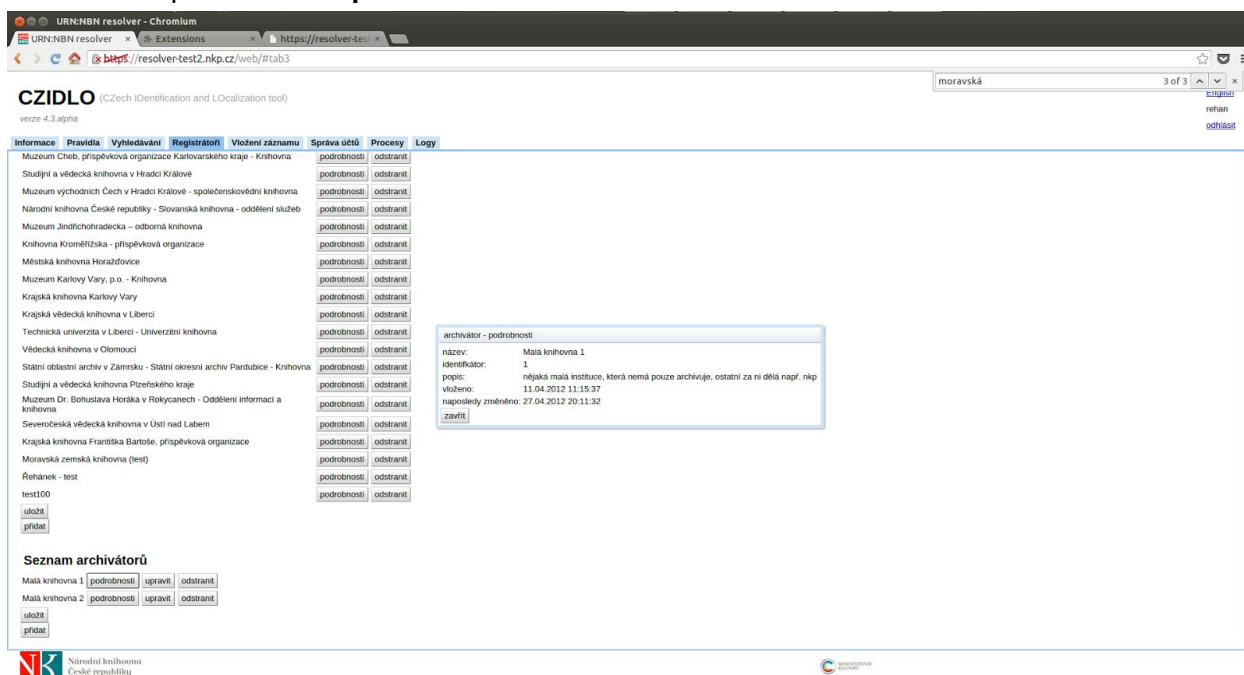
Identifikátor digitální knihovny lze také zjistit ze záznamu registrátora přes API:

<https://resolver-test2.nkp.cz/api/v3/registrars/tst02>.

- *archiverId* (*registrace DD*) - Nepovinný identifikátor archivátora, pokud sám registrátor není zároveň archivátorem. Lze zjistit v záložce **Registrátoři** dole, v sekci **Seznam**

OAI Adapter - uživatelský manuál

archivátorů po rozkliknutí **podrobnosti** u archivátora.



- **url (import DI)** - URL konkrétního digitálního dokumentu v digitální knihovně (tj. instance). To často bývá odvoditelné z identifikátoru, který je obsažen v metadatovém záznamu. Např. pro Kramerius se tvoří takto: <http://kramerius.nkp.cz/search/handle/> + pid (perzistentí identifikátor).
- **titleInfo (registrace DD)** - Tento element obsahuje různé názvové údaje, které se liší podle typu dokumentu. Třeba monografie má povinný element *title* a nepovinný *subTitle*. Ale číslo periodika (periodicalIssue) musí mít hned tři povinné názvové údaje - *periodicalTitle* (název celého periodika), *volumeTitle* (název ročníku), *issueTitle* (název čísla). Tenhle fakt musí OAI Repozitář zohlednit a poskytovat dostatečně popisná metadata.

Report.txt a logy

Výstupem OAI Adapteru je soubor report.txt. Ten obsahuje popis průběhu pro každý jednotlivý záznam sklizený přes OAI, dále transformovaný a použitý pro operace CZIDLO API. Na konci reportu je souhrn zpracovaných záznamů a počtů záznamů v jednotlivých výsledných stavech. Příklad reportu: <https://github.com/NLCR/CZIDLO/blob/master/examples/report.txt>.

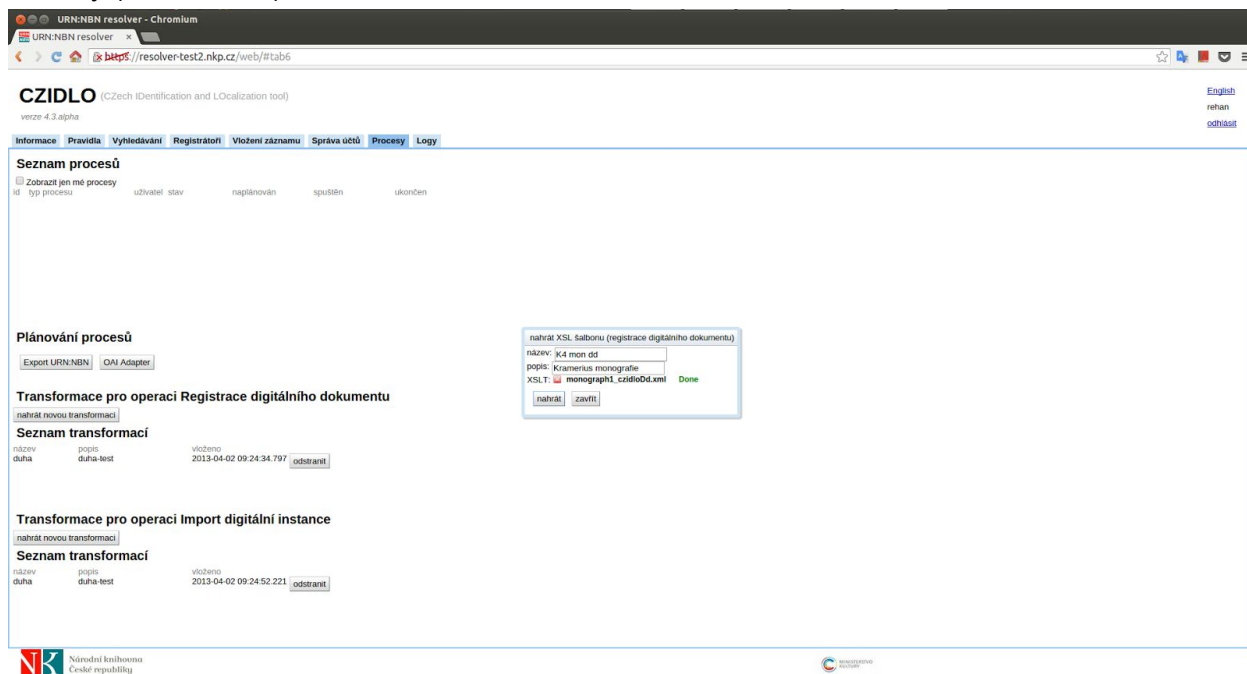
Ostatní události OAI Adapter ukládá do logu. Jde zejména o varování a chyby, které způsobí předčasné ukončení celého průběhu OAI Adapter. Při spuštění procesu přes webové rozhraní je log k dispozici ke stažení, stejně jako report. Při spuštění procesu z terminálu/příkazové řádky je log vypisován na standardní výstup.

OAI Adapter - uživatelský manuál

Spuštění přes webové rozhraní

Nejednodušší způsob použití OAI Adapteru je jeho spouštění jako procesu na serveru přes webové rozhraní. To je dostupné na adrese <http://resolver.nkp.cz>. Pro práci s procesy je nutné se přihlásit (v pravém horním rohu).

Před samotným spuštěním procesu je nutné nahrát šablony pro transformaci výstupu OAI do vstupu pro operace registrace DD a import DI. To provedete kliknutím na **nahrát novou transformaci** v sekcích **Transformace pro operaci Registrace digitálního dokumentu** a **Transformace pro operaci Import digitální instance** v záložce **Procesy**. V následujícím dialogu je potřeba vyplnit název a popis transformace a nahrát samotný soubor transformační šablony (soubor .xslt).



Nahrané transformační šablony jsou uloženy na serveru a spojeny s přihlášeným uživatelem. Pro další spuštění procesů je již nebude nutné znovu nahrávat.

Samotné spuštění procesu se provádí tlačítkem **OAI Adapter** v sekci **Plánování procesů** v záložce **Procesy**.

V dialogu pro konfiguraci procesu je třeba doplnit následující⁴:

- Zvolit kód registrátora. Pokud v seznamu není žádný registrátor, nebo chybí registrátor, s kterým jste měli v úmyslu pracovat, znamená to, že přihlášenému uživateli není přiřazeno právo spravovat registrátora. S požadavkem na přidělení práva se obraťte na kurátora/administrátora.
- Vybrat mód registrace.
- Vyplnit OAI base URL.
- Vyplnit OAI metadatový prefix.
- Vyplnit OAI set. Toto pole není povinné.

⁴ Více v kapitole Vstupní parametry.

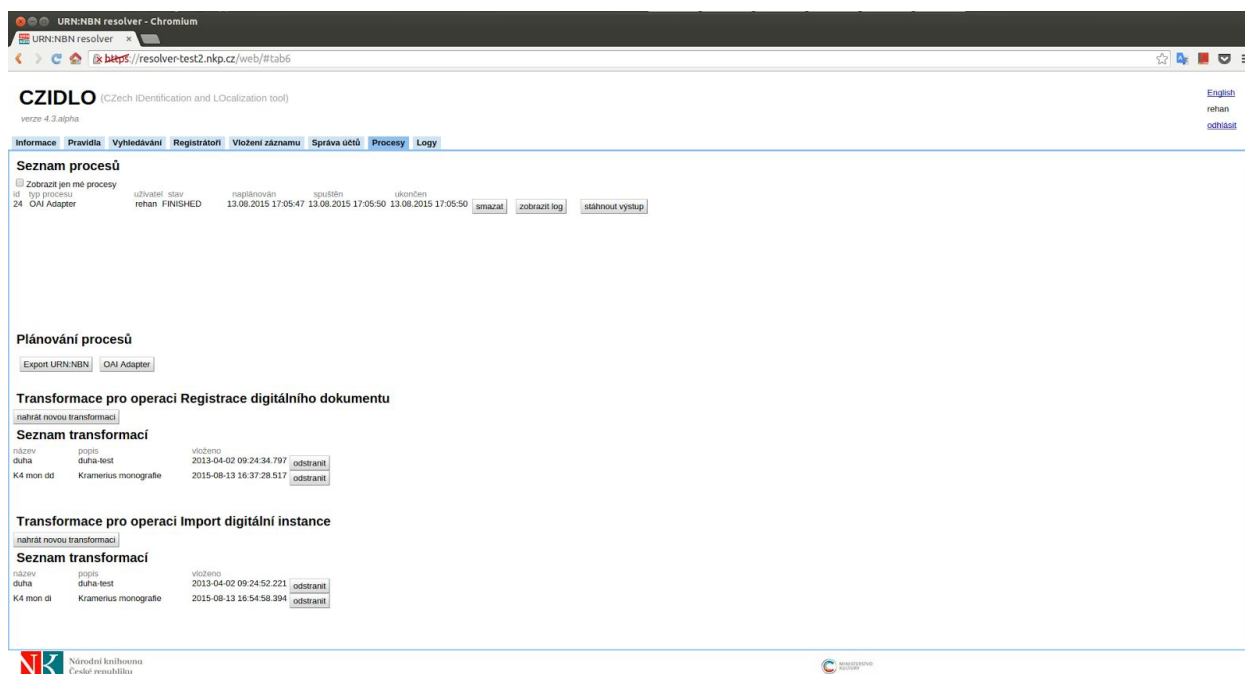
OAI Adapter - uživatelský manuál

- Zvolit transformační šablonu pro vstup operace registrace DD.
- Zvolit transformační šablonu pro vstup operace import DI.

Tlačítkem **Naplánovat proces** bude proces uložen a naplánován (stav SCHEDULED). Po nějaké době, která záleží na počtu a prioritě dalších naplánovaných procesů, bude proces spuštěn (RUNNING). Proces může být uživatelem předčasně ukončen (CANCELED). Nakonec proces skončí a dostane se buď do stavu KILLED (ukončen systémem), FAILED (chyba), nebo FINISHED (úspěšně dokončen).

Stavy procesů můžete průběžně kontrolovat v sekci **Seznam procesů** v záložce **Procesy**. Zobrazené procesy lze filtrovat jen na ty, které vlastní aktuálně přihlášený uživatel, a to zaškrtnutím **Zobrazit jen mé procesy**.

OAI Adapter - uživatelský manuál



Tlačítkem **zobrazit log** se otevře nové okno prohlížeče obsahující logy o průběhu/chybě procesu. Tlačítko je přítomno u probíhajících a dokončených procesů. Probíhající procesy je možné zrušit tlačítkem **zrušit**.

U úspěšně dokončených procesů (FINISHED) je dále zobrazeno tlačítko **stáhnout výstup**, které umožní stáhnout soubor, jenž je výstupem procesu, do vašeho počítače. U procesu OAI Adapter je výstupem soubor *report.txt*.

Záznamy všech dokončených procesů (FINISHED, FAILED, CANCELED, KILLED) lze smazat tlačítkem **smazat**.

Spouštění z příkazové řádky

Instalační balík CZIDLO⁵ obsahuje soubor *oaiAdapter-VERZE.jar*. Ten je určen pro lokální běh OAI Adapteru. Pro jeho spouštění je nutné mít nainstalováno prostředí Java Runtime Environment (JRE), alespoň ve verzi 6 (1.6). To obsahuje program java, který spouští kód v balíku jar. Aktuální verze JRE 8 je k dispozici zde:

<http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>

Spuštění OAI Adapteru vypadá následovně (Linux):

```
java -jar oaiAdapter-4.3.jar /path/to/configuration.properties
```

pro Windows:

```
java -jar oaiAdapter-4.3.jar C:\path\to\configuration.properties
```

Na výstup jsou vypisovány některé chybové a informační hlášky, proto může být vhodné výstup přesměrovat do souboru:

⁵ Instalační balíky jsou dostupné zde: <https://github.com/NLCR/CZIDLO/releases>.

OAI Adapter - uživatelský manuál

```
java -jar oaiAdapter-4.3.jar /path/to/configuration.properties >/path/to/logs.txt &
```

Configuration.properties

Soubor *configuration.properties* obsahuje samotné parametry běhu⁶. Jeho obsah vypadá nějak takto:

```
#####
# OAI PROVIDER #
#####
oaiAdapter.oai.baseUrl=http://duha-devel.mzk.cz/oai
oaiAdapter.oai.metadataPrefix=oai_dc
#@optional
oaiAdapter.oai.setSpec=monographs

#####
# CZIDLO API #
#####
oaiAdapter.czidlo_api.baseUrl=resolver.nkp.cz/api
oaiAdapter.czidlo_api.login=czidlo_user_123
oaiAdapter.czidlo_api.password=*****
oaiAdapter.czidlo_api.registrarCode=aba001
oaiAdapter.czidlo_api.registrationMode=BY_REGISTRAR
oaiAdapter.czidlo_api.ignoreInvalidCertificate=false

#####
# XSL TRANSFORMATIONS #
#####
oaiAdapter.digDocRegistrationXsl=/path/to/oaiDc_to_czidloDd.xsl
oaiAdapter.digInstImportXsl=/path/to/monograph_oaiDc_to_czidloDi.xsl

#####
# XSD FOR VALIDATING CZIDLO API INPUTS #
#####
oaiAdapter.digDocRegistrationXsdUrl=http://resolver.nkp.cz/api/v3/digDocRegistration.xsd
oaiAdapter.digInstImportXsdUrl=http://resolver.nkp.cz/api/v3/digInstImport.xsd

#####
# REPORT FILE #
#####
oaiAdapter.reportFile=/path/to/report.txt
```

⁶ Více v kapitole Vstupní parametry.

OAI Adapter - uživatelský manuál

Kromě základních parametrů popsanych v sekci *Vstupní parametry*, jsou zde ještě navíc parametry *oaiAdapter.czidlo_api.**, které slouží k napojení OAI Adapteru na CZIDLO API:

- **oaiAdapter.czidlo_api.baseUrl** - Základní URL CZIDLO API bez uvedení protokolu. Protokolem je vždy https, nicméně jej neuvádějte. Na konci by nemělo být lomítko.
- **oaiAdapter.czidlo_api.login** - Uživatelské jméno (login) uživatele, který má přiřazena práva pro daného registrátora.
- **oaiAdapter.czidlo_api.password** - Heslo uživatele.
- **oaiAdapter.czidlo_api.registrarCode** - Kód registrátora.
- **oaiAdapter.czidlo_api.registrationMode** - Mód registrace.
- **oaiAdapter.czidlo_api.ignoreInvalidCertificate** - Zda má být ignorován neplatný serverový certifikát při operacích API. Určeno jen pro testování, v produkčním provozu vždy uvádějte *false*.

O tomto dokumentu

Verze: 1.1

Vytvořeno: 12. 7. 2015 (Martin Řehánek)

Naposledy upraveno: 29. 7. 2015 (Martin Řehánek)