



národní
úložiště
šedé
literatury

Manuál k lokální instalaci CDS Invenia a rámci spolupráce s Národním úložištěm šedé literatury (NUŠL) Národní technické knihovny (NTK)

Frantíková, Bohdana
2010

Dostupný z <http://www.nusl.cz/ntk/nusl-111905>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte autora-Zachovejte licenci 3.0 Česko

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 27.09.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Manuál k lokální instalaci CDS Invenia a rámci spolupráce s Národním úložištěm šedé literatury (NUŠL) Národní technické knihovny (NTK)

1.0 Verze

Obsah

Instalace Invenia

1. Lokální instalace defaultního CDS Invenia

1.1. Základní poznámky k užívání VirtualBoxu

1.2. Běží virtuální počítač (exportovaný do formátu OVF) ještě na jiném virtuálním prostředí?

2. Poznámky ke konfiguraci CDS Invenia

2.1. Poznámky k síťovému nastavení CDS Invenia

2.2. Protokoly a certifikáty

2.3. Konfigurační soubor "invenio-local.conf"

Spravování Invenia

3. Základy práce s CDS Inveniem

3.1. Popis a vyjmenování základních periodických úkonů

3.2. Organizace archivu podle účelu

3.3. Formáty digitálních objektů

4. Přehled modulů CDS Invenia

4.1. Moduly pro získávání a zpracování záznamů

4.2. Moduly pro další zpracování a kontrolu dat

4.3. Moduly zabývající se personalizací uživatelských kont

4.4. Administrátorské moduly

4.5. CDS Invenio "workflow" diagram znázorňující spolupráci jednotlivých modulů

5. Databáze MySQL

6. Slovo závěrem

Instalace Invenia

Tento text popisuje jak nainstalovat, nakonfigurovat a spravovat defaultně nakonfigurovaný systém CDS Invenio. Tento software si spolupracující organizace nainstaluje na vlastním hardwaru a dále upraví dle vlastních požadavků při zachování odpovídajících struktur (sbírek, formulářů, ..) určených Národním úložištěm šedé literatury (NUŠL). Toto omezení se ovšem týká pouze šedé literatury, a tedy správa ostatních digitální objektů na lokálním úložišti je plně v režii dané organizace.

1. Lokální instalace defaultního CDS Invenia

Běžně zájemce o provozování CDS Invenia musí vytvořit balíček (make) z volně dostupného zdrojového kódu, ten pak instalovat (install) a provést primární konfiguraci (<http://invenio-software.org/wiki/Installation/Download>). Instalace CDS Invenia také předpokládá instalaci dalšího software a jeho nastavení: MySQL server, Apache server, Python atd. Ovšem NTK připravila pro spolupracující organizace "přednastavenou" defaultní NUŠL verzi, která předpokládá pouze instalaci aplikace Virtual Box. Tato aplikace vytvoří na lokálním serveru virtuální prostředí, které může obsahovat jeden či více virtuálních počítačů. Na jednom z nich poběží CDS Invenio nad operačním systémem Linux-Debian.

Navíc po instalaci Virtual Boxu není třeba Debian ani CDS Invenio instalovat, ale celý virtuální počítač se importuje ve Virtual Boxu, a to již s přednastavenou konfigurací (pro potřeby NUŠLu) a po nastavení síťových parametrů virtuálního stroje může běžet. Cílem instalace Virtual Boxu namísto přímo CDS Invenia bylo vyhnout se závislosti na operačním systému lokálního serveru, protože VirtualBox lze nainstalovat na všechny víceméně známé platformy.

Celý postup v bodech:

1. Instalace lokálního VirtualBoxu (<http://www.virtualbox.org/wiki/Downloads>)
2. Import vyexportovaného virtuálního počítače i s CDS Inveniem do lokálního VirtualBoxu. Heslo pro uživatele „invenio“ je „invenio“, heslo pro „root“ je „invenio“
3. Úprava síťových parametrů (IP, hostname, atd.) virtuálního stroje
4. Editovat a nastavit především první část konfiguračního souboru `/opt/cds-invenio/etc/invenio-local.conf`
5. Aktualizace nastavení systému z příkazové řádky: `inveniocfg --update-all`
6. Vygenerování konfiguračních souborů pro apache: `inveniocfg --create-apache-conf`
7. restartovat apache: `/etc/init.d/apache2 restart`

Po importu virtuálního počítače je ještě před jeho spuštěním nezbytné zkontrolovat nastavení. To se může v důsledku odlišných OS a verzí Virtual Boxu při importu mírně změnit.

Zde je nastavení virtuálního počítače před exportem:

1. Typ operačního systému: Linux Debian
2. Operační paměť: 512 MB
3. Video paměť: 64 MB
4. ACPI: Povoleno
5. IO ACPI: Zakázáno
6. VT-x/AMD-V: Povoleno
7. Přímý přístup k hardware: Zakázán
8. PAE/NX: Zakázáno
9. 3D akcelerace: Povoleno
10. Pevné disky: SATA1 (invenio.vmdk)
11. Zvuk: Zakázán
12. Síť: Karta 1: Intel PRO/1000 MT Desktop (Karta použitá v síťovém mostu, e1000g0 - Intel PRO/1000 Gigabit Ethernet)
13. Sériové porty: Zakázány
14. USB: Zakázán
15. Sdílené složky: Žádné
16. Vzdálená plocha: Port VRDP serveru (3390/3389)

Důležité je především zkontrolovat nastavení síťové karty aby skutečně byla v síťovém mostu a to se síťovou kartou, přes kterou já váš server připojený do sítě.

1.1. Základní poznámky k užívání VirtualBoxu

Uvedeme několik příkazů, jimiž lze VirtualBox kontrolovat z příkazové řádky. Grafickou konzoli otevřeme příkazem:

```
VirtualBox
```

Jak na restart: Po připojení se na server jako uživatel „ohleduplně“ zastavíme virtuální počítač příkazem:

```
VBoxManage controlvm invenio acpipowerbutton
```

kde invenio.ntkcz.cz je název virtuálního počítače. Pokud i víc než po jedné minutě hlásí příkaz

```
VBoxManage showvminfo invenio | grep Status
```

odpověď "State: Running...", vypneme VirtualBox „natvrdo“:

```
VBoxManage controlvm invenio poweroff
```

Nový start:

```
VBoxManage startvm invenio --type vrdp
```

kde vrdp znamená Virtual Remote Desktop Protocol – počítač se spustí na pozadí a lze s k němu připojit přes ssh. Je potřeba mít povolenou vzálenou plochu a mít nastavení volný port pro VRDP v konfiguraci virtuálního počítače. Bez tohoto parametru se vám otevře grafické okno. Více lze zjistit z manuálů VirtualBoxu pro koncové uživatele:

<http://download.virtualbox.org/virtualbox/3.1.0/UserManual.pdf>

<http://www.virtualbox.org/manual/UserManual.html>

nebo přímo z příkazové řádky

```
VBoxManage -h
```

1.2. Běží virtuální počítač (exportovaný do formátu OVF) ještě na jiném virtuálním prostředí?

Virtuální počítač s CDS Inveniem je zřejmě možné převést i do jiného virtuálního prostředí než do prostředí VirtualBox. Vyzkoušeno je též prostředí VMWare ESXi (další často používané virtuální prostředí).

2. Poznámky ke konfiguraci CDS Invenia

NTK dodává “předkonfigurované” CDS Invenio pro potřeby šedé literatury. Organizace si však může svůj lokální repozitář dále upravit (ovšem bez narušení sbírek, formulářů, atd.) dle místních požadavků. Zde jsou uvedeny podstatné poznámky a doporučení pro další konfiguraci.

2.1. Poznámky k síťovému nastavení CDS Invenia

Konfigurační soubory pro apache se generují automaticky pomocí výše zmíněného souboru

```
/opt/cds-invenio/bin/inveniocfg --create-apache-conf
```

a ukládají se do adresáře

```
/opt/cds-invenio/etc/apache
```

V adresáři konfigurace apache

```
/opt/apache2/sites-enabled a /opt/apache2/sites-available
```

jsou pak jen symbolické odkazy na konfigurační soubory uložené v Inveniu.

2.2. Protokoly a certifikáty

Invenio vyžaduje certifikovaný protokol https. Z tohoto důvodu je přednastavené a virtualizované CDS Invenio primárně samocertifikováno (self-signed certificate). Při publikaci CDS Invenia do vnější internetové sítě se tedy předpokládá získání veřejného certifikátu od nějaké autority.

Bezpečnostní klíče v Apache: adresář klíčů pro certifikát je

```
/etc/apache/ssl
```

2.3. Konfigurační soubor "invenio-local.conf"

Velmi podstatný krok v konfiguraci CDS Invenia představuje editace konfiguračního souboru invenio-local.conf.

CDS Invenio obsahuje dva konfigurační soubory: invenio.conf a invenio-local.conf, přičemž první z nich obsahuje veškeré možné parametry (je vhodné ho nechat beze změny), zatímco druhý je určen k editaci a použití pro vlastní chod systému („překrývá“ ten první).

Po jakékoliv změně konfiguračního souboru je nutné aktualizovat tuto změnu spuštěním příkazu

```
/opt/cds-invenio/bin/inveniocfg --update-all
```

Příklady editace:

Nastavení domén: CFG_SITE_URL, CFG_SITE_SECURE_URL, ..

Vypnutí neužívaných jazyků: Zakomentování odpovídající řádky v souboru invenio-local.conf.

Název lokálního úložiště: CFG_SITE_NAME

Upozornění na chybu či důležitou akci: CFG-SITE-ADMIN-MAIL

Spravování Invenia

3. Základy práce s CDS Inveniem

3.1. Popis a vyjmenování základních periodických úkonů

Mnoho procesů nad daty archivu může být spuštěno v periodickém módu. Chod periodických (i neperiodických) procesů zajišťuje démon BibSched, který kontroluje frontu úkolů, spouští jednotlivé procesy v naplánovaném pořadí. Hlavní odkaz: <http://invenio-demo.cern.ch/help/admin/howto>

Například, proces zajišťující pravidelné indexování nových dokumentů spustíme příkazem (a bude se opakovat každou hodinu)

```
bibindex -f50000 -slh
```

Proces zajišťující formátování metadatových záznamů je spuštěn příkazem (a opakován každou hodinu)

```
bibreformat -slh
```

Proces pro update cache pro sbírky je spuštěn příkazem (opět každou hodinu)

```
webcoll -slh
```

Proces, který prochází obsahy záznamů a na základě určitých kritérií (implicitně dle podobnosti slov) vyhodnocuje, které záznamy jsou si podobné (zpřístupňuje funkci hledání podobných záznamů), spouští se každou hodinu.

```
Bibrank -slh
```

Proces harvestování definovaných archivů se provede pomocí démona (každých 24 hodin)

```
oaiharvest -s 24h
```

Vystavení archivu pro harvestování jinými organizacemi (aktualizace jednou za 24 hodin, parametr global popisuje vystavený set archivu):

```
oaiarchive -apo global -s24h
```

Vystavený archiv invenio.ntkcz.cz je k nalezení na <http://invenio.ntkcz.cz/oai2d?verb=ListRecords&set=global&metadataPrefix=marcxml>

3.2. Organizace archivu podle účelu

Celá instalace:

```
/opt/cds-invenio
```

Konfigurační soubor CDS Invenia, jeho modulů včetně transformačních souborů:

```
/opt/cds-invenio/etc
```

Jazykové mutace:

```
/opt/cds-invenio/share
```

Cache, uložené soubory, logy, temp:

```
/opt/cds-invenio/var
```

Samotné digitální objekty (fulltexty, multimedia) jsou uloženy v adresáři:

```
/opt/cds-invenio/var/data/files
```

Každému dokumentu je systémem přiděleno identifikační číslo, které určuje adresář, ve kterém je tento dokument uložen. Například, soubory odpovídající dokumentu s identifikačním číslem 14 budou uloženy v adresáři:

```
/opt/cds-invenio/var/data/files/g0/14
```

Počet dokumentů v podadresářích g0 (g1, ..), které obsahují skupiny dokumentů, je možné nastavit v parametru

CFG-FILE-DIR-SIZE

v souboru `invenio-local.conf`. Jednomu dokumentu může odpovídat více jednotlivých souborů: různých formátů nebo různých verzí.

Na každou podsbírku připadá parametr, ve kterém se ukládá počet momentálně uložených dokumentů – adresář:

```
/opt/cds-invenio/var/data/submit/counters
```

Všechny dokončené či probíhající procesy vkládání jsou v adresáři:

```
/opt/cds-invenio/var/data/submit/storage
```

Většina výkonných souborů se nachází v pythoní knihovně

```
/opt/cds-invenio/lib/python/invenio
```

3.3. Formáty digitálních objektů

CDS Invenio umožňuje vkládat libovolné formáty digitálních objektů. Přesto pro snadnou organizaci, přístupnost a migraci (při zastarávání formátů) je vhodné zúžit množinu přijímaných formátů (nebo alespoň doporučit uživatelům). Například zvolit pro každý typ digitálních objektů dva či tři preferované formáty

Podporované formáty NUŠL pro jednotlivé typy digitálních objektů budou definovány v rámci dohody o spolupráci.

V Inveniu se formáty zachovávají, tedy. netransformují se na jiné (standardní) formáty. A není tedy možné (bez pomoci jiné aplikace) při importu automatizovaně převádět jiné formáty na formát standardní. Pokud je vložen dokument v pdf formátu CDS Invenio si jej může založit ještě jako komprimovaný postscript, tj. `*.ps.gz`.

4. Přehled modulů CDS Invenia

V této kapitole jsou popsány základní nástroje konfigurace CDS Invenia, případně odkazy na webové zdroje pro detailnější popis. CDS Invenio se skládá z několika modulů, přičemž každý tento modul je vybaven specifickou funkcionalitou. Tyto moduly jsou stručně popsány s odkazy na detailnější popisy a znázorněny workflow diagramem v následujících podkapitolách (viz <http://invenio-demo.cern.ch/help/admin/>).

4.1. Moduly pro získávání a zpracování záznamů

Metadatové záznamy získány dvěma způsoby: Buď zápisem do příkazové řádky (název modulu plus parametry). Příkladem jsou moduly `BibConvert`, pomocí něhož se konvertují metadata do vnitřního formátu CDS Invenia (MARCXML), a `BibUpload`, pomocí něhož se nahrají MARCXML soubory do CDS Invenio. A nebo užitím interaktivního webového rozhraní: Příkladem jsou moduly `WebSubmit`, pomocí něhož definujeme vše co souvisí s vkládáním metadat, popř. samotných dokumentů, `BibHarvest`, pomocí něhož se harvestují metadata z jiných OAI repozitářů, a `BibEdit`, pomocí něhož lze nahraná metadata modifikovat.

Modul	Popis	Rozhraní
BibHarvest	V příslušném rozhraní definujeme parametry pro harvestování jednotlivých repozitářů: jméno repozitáře, periodicitu harvestování, způsob konverze takto získaných záznamů, uploadování konvertovaných záznamů do CDS Invenia.	webové rozhraní
BibConvert	Konvertuje metadatové záznamy do zadaného formátu. Vhodné například v případech konvertování metadat ze starého systému či při konvertování metadat, které neodpovídají protokolu OAI.	příkazová řádka
BibUpload	Pomocí modulu, lze nahrát jednotlivé metadatové záznamy.	příkazová řádka
WebSubmit	V tomto rozhraní konfigurujeme formuláře a šablony pro vkládání záznamů odpovídajících různým sbírkám. Dokumenty mohou podléhat také peer-review či schvalování po vytvoření odpovídajícího formuláře pro danou sbírku.	webové rozhraní
ElmSubmit	Umožňuje konfiguraci vkládání záznamů emailem.	příkazová řádka
BibEdit	Díky tomuto modulu můžeme přímo modifikovat jednotlivé metadatové záznamy, mazat je či resetovat je.	webové rozhraní

Odkazy na detailnější popis těchto modulů

<http://invenio-demo.cern.ch/help/admin/bibindex-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibrank-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibclassify-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibformat-admin-guide>

<http://invenio-demo.cern.ch/help/admin/websearch-admin-guide>

<http://invenio-demo.cern.ch/help/admin/webstat-admin-guide>

4.2. Moduly pro další zpracování a kontrolu dat

K dalšímu "jemnějšímu" zpracování dat jsou k dispozici např. tyto moduly: BibIndex, který indexuje metadata, BibRank, pomocí něhož jsou metadata tříděna, BibFormat, který metadata formátuje pro výstup podle předepsaného vzoru, a WebSearch, pomocí něhož upravujeme rozhraní pro vyhledávání a vyhledávání samotné.

Modul	Popis	Rozhraní
BibIndex	Umožňuje konfigurovat "word files", definuje tedy která metadatová pole jsou indexována ke kterým slovním indexům. Tyto slovní indexy jsou pak užity při vyhledávání. Např. definujeme, že index "autor" je přiřazen MARCovými poli \$100 a \$700.	webové rozhraní

BibRank	Umožňuje konfigurovat různé třídící metody, které mohou být užity při vyhledávání: Impact Factor, ..	webové rozhraní
BibClassify	Umožňuje automaticky vyhledávat klíčová slova v dokumentech pro zvolené řízené slovníky a hesláře.	příkazová řádka
BibFormat	Tento modul zajišťuje prezentaci metadat, tedy definujeme pomocí něho různé metadatové formáty, ve kterých chceme mít metadata zobrazeny.	webové rozhraní
WebSearch	Umožňuje konfiguraci rozhraní pro vyhledávání ve sbírkách. Tyto sbírky zde definujeme, uspořádáme do dvou-úrovňového stromu. Dále definujeme indexy a další možnosti pro vyhledávání atd.	webové rozhraní
WebStat	Umožňuje konfiguraci statistik, ukazujících využitelnost jednotlivých záznamů.	příkazová řádka

Odkazy na detailnější popis těchto modulů

<http://invenio-demo.cern.ch/help/admin/bibindex-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibrank-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibclassify-admin-guide>

<http://invenio-demo.cern.ch/help/admin/bibformat-admin-guide>

<http://invenio-demo.cern.ch/help/admin/websearch-admin-guide>

<http://invenio-demo.cern.ch/help/admin/webstat-admin-guide>

4.3. Moduly zabývající se personalizací uživatelských kont

Moduly umožňující personalizaci rozhraní CDS Invenia různým požadavkům koncových uživatelů jsou např. tyto: WebSession, pomocí něhož identifikujeme uživatele a jejich osobní konfiguraci, WebBasket, který umožňuje manipulaci s daty formou "nákupních" košíků, a WebAlert, pomocí něhož si jednotlivý uživatel definuje upozornění zasílaná emailem.

Modul	Popis	Rozhraní
WebSession	Umožňuje kontrolu jednotlivých uživatelů a jejich konfigurace, případně hostů.	příkazová řádka
WebBasket	Umožňuje kontrolu a manipulaci s množinami uživatelských košíků.	příkazová řádka
WebAlert	Umožňuje kontrolu, konfiguraci a správu upozornění definovaných uživateli.	příkazová řádka
WebComment	Umožňuje manipulaci s komentáři a recenzemi přidanými uživateli či hosty.	webové rozhraní
WebMessage	Umožňuje konfiguraci zasílání a vzkazů.	příkazová řádka

Odkazy na detailnější popis těchto modulů

<http://invenio-demo.cern.ch/help/admin/websession-admin-guide>

<http://invenio-demo.cern.ch/help/admin/webbasket-admin-guide>

<http://invenio-demo.cern.ch/help/admin/webalert-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webcomment-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webmessage-admin-guide>

4.4. Administrátorské moduly

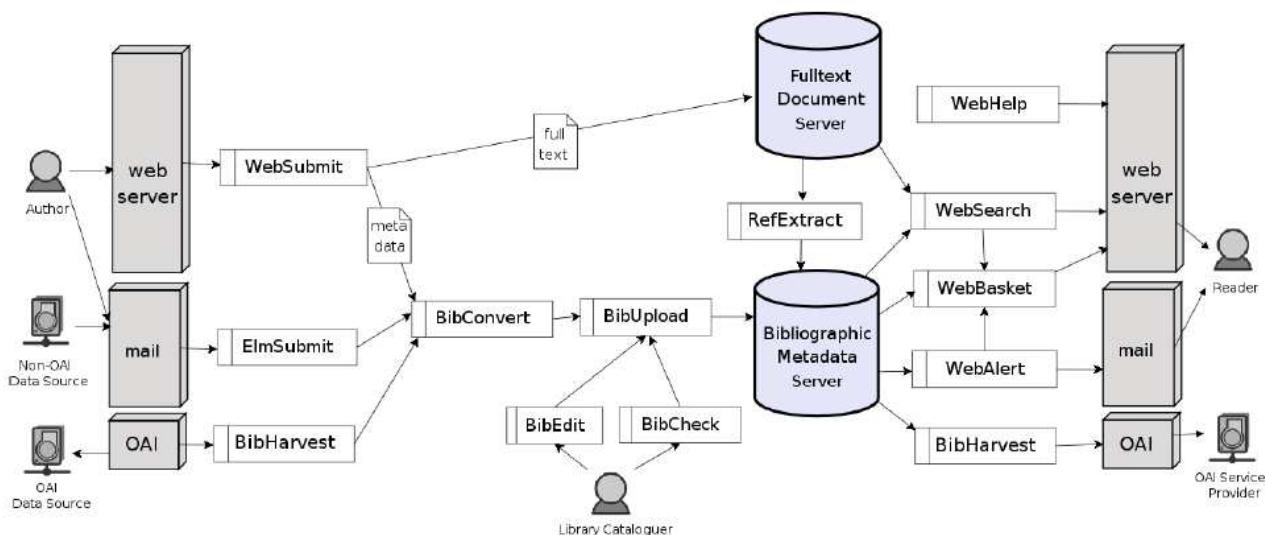
Moduly využívající ostatní moduly a zajišťující chod a vzhled CDS Invenia: BibSched, pomocí něhož manipulujeme a plánujeme úkoly pro další moduly, WebAccess, pomocí něhož definujeme přístup a práva k jednotlivým částem CDS Invenia pro dílčí správce, a WebStyle, pomocí něhož definujeme vzhled webu CDS Invenio.

Modul	Popis	Rozhraní
BibSched	Umožňuje kontrolu a plánování úkolů vykonávaných dalšími moduly, přidělování priorit odpovídajícím procesům, periodické spouštění atd.	příkazová řádka
WebAccess	Umožňuje definovat přístupová práva k jednotlivým administrátorským či knihovnickým činnostem. Např. správa jednotlivých modulů, vkládání dokumentů, schvalování a recenzování dokumentů atd.	webové rozhraní
WebStyle	Umožňuje přizpůsobit styl webového rozhraní CDS Invenio a CSS style sheet.	

Odkazy na detailnější popis těchto modulů

<http://invenio-demo.cern.ch/help/admin/websession-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webbasket-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webalert-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webcomment-admin-guide>
<http://invenio-demo.cern.ch/help/admin/webmessage-admin-guide>

4.5. CDS Invenio "workflow" diagram znázorňující spolupráci jednotlivých modulů



5. Databáze MySQL

CDS Invenio používá MySQL databázi. Databáze, která se jmenuje cds-invenio, obsahuje veškerá data související se záznamy uspořádaná do odpovídajících tabulek. K nahlížení a uvážené editaci databáze je v této instalaci připraveno webové rozhraní PhpMyAdmin, přístupné na adrese <http://www.your.site.cz/phpmyadmin>

Změna hesla MySQL musí být doprovázena přepsáním parametru CFG_DATABASE_PASS v konfiguračním souboru

```
/opt/cds-invenio/etc/invenio-local.conf
```

6. Slovo závěrem

Jedná se o jednu z prvních verzí tohoto manuálu. Přestože cílem bylo vypracovat manuál co nejsrozumitelněji, předpokládáme připomínky a dotazy, které laskavě zasílejte správci CDS Invenia v NTK na e-mail admin.invenio@techlib.cz.