

# Co je (staro)nového v DSpace

Vlastimil Krejčíř, [krejcir@ics.muni.cz](mailto:krejcir@ics.muni.cz)

Ústav výpočetní techniky, Masarykova univerzita, Brno

CZDSUG 2011, Ostrava

- Delegování práv.
- Autentizace přes IP adresy.
- Omezení viditelnosti, skrytí metadat.
- Export (CSV).
- Rozšířená konfigurace procházení Items (Multiple display).
- Podpora pro metadata v Google Scholar.
- Google Analytics (XMLUI) a Sitemap.
- Podpora pro linkování do SFX.

# Obsah přednášky II

- Embargo system.
- OpenSearch.
- OAI-ORE.
- SWORD.
- Statistiky a facety (Apache SOLR).
- Další vylepšení a nové funkce.

# Delegování administrátorských práv

- systém delegování a propagace administrátorských práv
- požadován už od verze 1.2!
- administrátor pro Community a Collection
- oprávnění se dědí na nižší objekty
- *Edit* → *Community Administrators*:
  
- funguje přes systém skupin (Group)
- podrobné nastavení atributů, které se dědí v konfiguračním souboru
  - `core.authorization.community-admin.*`
  - `core.authorization.collection-admin.*`

# Autentizace přes IP adresy

- není přímou autentizací
- uživatel se musí přihlásit přes některou jinou metodu
- na základě své IP adresy se stává členem vybraných skupin
- podpora IPv6
- různé zápisy rozsahů
  - `plugin.sequence.org.dspace.authenticate.AuthenticationMethod = org.dspace.authenticate.IPAuthentication`
  - `authentication.ip.<skupina> = 147.251.0.0/16, \`  
`-147.251.17.170`

# Omezení viditelnosti

- OAI-PMH, RSS i subscriptions nekontrolují přístupová práva k Item
- informace o Items nedostupných bez autorizace jsou k dispozici
- lze však nastavit omezení na Items s přístupem pro anonymní uživatele
- snižuje výkon (zejména OAI serveru) – ListIdentifiers pak spouští kontrolu práv
  - `harvest.includerestricted.*`

- je možné vybraná metadatová pole skrýt před uživatelem
- dostupná pouze pro administrátora
- platí pro web, OAI i RSS
  - `metadata.hide.<schema>.*`

# Rozšířená konfigurace procházení obsahem DSpace

- systém procházení obsahu repozitáře doznal od verze 1.4 značných změn
- je kompletně konfigurovatelný
- možnost budování vlastních indexů
- konfigurovatelnost je rozsáhlá
- způsob třídění, počty záznamů na stránce, ...
  - `webui.browse.*`



- Google Scholar hledá v HTML hlavičce metadata ve specifickém formátu
- DSpace 1.7 podporuje automatickou tvorbu těchto hlaviček (JSPUI)
- možnost definovat vlastní mapování
  - `google-metadata.enable = true`

# Google Analytics a Sitemap

- při používání XMLUI (dříve Manakin) možnost napojení Google Analytics
- `xmlui.google.analytics.key`
- možnost tvorby Sitemapy pro vyhledávače
- `sitemap.dir`
- `sitemap.engineurls`

- DSpace je možné napojit na SFX server (OpenURL)
- možné pouze pevné napojení na jeden server
- na základě konfiguračního souboru se generuje OpenURL
- konfigurace je flexibilní (definuje jaká metadata do OpenURL zařadit)
  - `sfx.server.url`

# Embargo System

- Embargo System realizuje princip *moving wall*
- přístup k některým Items může být časově omezen (např. objekt je přístupný veřejnosti až 90 dnů po vložení do DSpace)
- možnost omezit intervalem (od vložení) i přímo datem, ke kterému má být Item zveřejněna
- informace o embargu přímo v metadatech
- výchozí java handlers (*Setter* a *Lifter*) – nastavení práv, možno přepsat
- řízená kontrola přes *crontab*
  - `embargo.field.*`
  - `embargo.terms.*`
  - `plugin.single.org.dspace.embargo.*`

- technologie na unifikaci vyhledávání a zpracování výsledků (obvykle strojově) – podobná SRW/U
- obvyklá u webů, které mají vyhledávání integrováno
- hledání v DSpace lze pomocí ní integrovat do aplikací třetích stran (např. do webového prohlížeče)
- zatím jen základní implementace
  
- na URL `open-search/?query=<your query>` (REST)
- formát pro zasílání výsledků je Atom, RSS resp. i HTML
  - `websvc.opensearch.*`

- protokol pro popis a agregaci digitálních objektů v prostředí webu
- provider i harvester – úzce napojen na OAI-PMH
- harvester – zvyšuje interoperabilitu mezi repozitáři, "vzdálené kolekce" (klikatelně přes uživatelské rozhraní)
- problémový Dublin Core?
  
- viz *Harvesting Settings* při editaci kolekce
- RDF a Atom
- vše řešeno přes Crosswalks (Java pluginy)
- nastavení harvesteru pro periodické sklizení
  - `harvester.*`
  - `ore.*`

# SWORD

## Import do DSpace

- standard definující jednotný formát pro popis digitálního objektu
- SWORD objekt může být importován do DSpace
- implementují i jiné repozitáře (mj. Fedora a EPrints)
- externí klienti (obvykle přes web)
  - <http://client.swordapp.org/client/>
- klient přímo pro DSpace?
  
- interně pro DSpace .zip s metadaty a soubory
- METS přes crosswalk jde do DIM a následně do DSpace
- sword/servicedocument
- sword/deposit/<handle>
  - sword.\*

# Statistiky (Apache SOLR)

- podrobný modul pro statistiky je nyní integrován přímo do DSpace
- umožňuje podrobnou konfiguraci sledování přístupů apod.
- dostupný přímo přes webové rozhraní
- přednastavené výsledky typu "10 nej"
- postaven na Apache SOLR (rozšiřuje možnosti Lucene) – samostatný server
- analýza logů
  - solr.\*



# Facety (Apache SOLR)

- SOLR poskytuje nad Lucene tzv. Faceted search (od verze 1.7)
- uživatel nemusí hledat komplikovaným dotazem ("Borůvka" AND "Linear algebra")
- zadá dotaz ("Borůvka") a následně jej ve výsledcích upřesňuje (filtruje)
- podobně pracuje např. Amazon
  
- vlastní konfigurační soubor `dspace-solr-search.cfg`
- lze nastavit, z jakých metadatových polí se facety mají tvořit
- SOLR jako samostatný server
  - `solr.search.*`

# Další vylepšení a nové funkce

- optimalizace kódu – browse, Lucene Indexing
- vícejazyčnost
- podpora MS PowerPoint při indexaci fulltextu
- vylepšená podpora XMLUI (Mirage theme)
- optimalizace SQL
- vylepšený import/export (včetně struktur)
- Recommend this item
- podpora pro authority (LoC)
- vnitřně Event system
- ...

DSpace 2...

Děkuji za pozornost.