

ResourceSync: synchronizace zdrojů na internetu



Vlastimil Krejčíř
krejcir@ics.muni.cz

22. května 2014

■ ResourceSync

Co je ResourceSync?

Protokol pro synchronizaci **jakýchkoli zdrojů** na internetu, který umožňuje **efektivní** a **flexibilní** synchronizaci nezávisle na typu zdroje a rychlosti jeho změny.

Standard ANSI/NISO Z39.99-2014

ResourceSync Framework Specification

<http://www.openarchives.org/rs/toc>

■ Poznámky na úvod

K přednášce

Vynecháme technické detaily – ty si lze přečíst v článku (a samozřejmě ve specifikaci). Budeme mírně zjednodušovat.

K článku

Všechno je jinak... aneb máme finální verzi ResourceSync 1.0.

■ ResourceSync: motivace

Proč nový protokol?

- nutnost synchronizace (zdrojů – dat)
- současné nástroje příliš „low level“
nutná těsná spolupráce
- synchronizace dat (OAI-PMH nestačí)
- udělat nový **moderní** protokol, který plní současné požadavky na pohodlnou synchronizaci

■ ResourceSync: pozadí, historie

Kdo za tím stojí?

Podporován organizacemi **NISO** a **OAI**.

Lidé

Mezinárodní pracovní skupina, mj. *Carl Lagoze* a *Herbert Van de Sompel* .

Historie

- první info v půlce roku 2012
- nyní verze 1.0
- **ANSI/NISO Z39.99** standard 21. května 2014

■ OAI-PMH vs. ResourceSync

Proč nám nestačí OAI-PMH?

- primárně synchronizace metadat
 - přenos mimo protokol OAI-PMH 😊
 - kódování do metadat v XML 😞
- opakované sklízení (pull-based)
 - není vždy synchronní 😞
- selektivní sklízení (Sets)
 - předem definované kolekce

■ Základní definice

Zdroj (Resource)

- to, co synchronizujeme (data, metadata, ...)

Poskytovatel (Source)

- dává zdroje k synchronizaci
- poskytuje další služby

Cíl (Destination)

- chce být synchronní, s poskytovatelem
- přebírá změněné zdroje

■ Funkce a vlastnosti

- základní identifikátor je URI
- inkrementální synchronizace
- hromadné synchronizace
- intervalové (čas) synchronizace
- synchronizace starších verzí dat (verzování)
- informování (notifikace) o změnách
- provázání synchronizovaného obsahu
- inzerování podpory protokolu
- **modularita**

■ Možnosti protokolu

Pohled poskytovatele

- informace o zdrojích a možnostech poskytovatele
- informace o změnách ve zdrojích
- synchronizace dle času (verzování, intervaly)
- hromadná synchronizace (v balíku ZIP)
- provázání (odkazy) na jiné zdroje
- discovery – dát o sobě vědět

■ Možnosti protokolu II

Pohled cíle

- **iniciální synchronizace**
poprvé chci vše
- **inkrementální synchronizace**
chci pouze změny
- **audit**
jsme synchronní?

■ Metadata zdroje a provázání zdrojů

Každý zdroj má technická metadata (URI, datum změny, ...) a **odkazy**:

- mirror
- alternativní zdroj
- rozdílové aktualizace (patch)
- kolekce (kde se zdroj nachází)
- metadata o zdroji (describedby)
- předchozí verze zdroje (funkce verzování)

■ Technické informace

- postaveno na typu *Sitemap*
- rozšířeno o nové elementy pro popis zdroje
 - nový *Namespace* `rs`
 - popis zdroje `<rs:md>`
 - odkaz na jiný zdroj `<rs:ln>`
- pečlivá a přesná práce s časem
protokol Memento

■ Notifikace: zjišťování změn

Pull-based

- opakované dotazy na změny
Co je nového?
- vznikají prodlevy
Jak často se ptát?
- nejsou změny \Rightarrow zbytečný dotaz

Push-based

- poskytovatel informuje o změnách
- notifikace pouze, když ke změně dojde
- cíl se přihlásí k odebírání notifikací

■ ResourceSync náhrada OAI-PMH

- ResourceSync dokáže nasimulovat OAI-PMH náhrada za všechny příkazy
 - identifikace (Identify)
 - výpis identifikátorů (ListRecords)
 - stažení záznamu/ů (GetRecord, ListRecords)
- funkce navíc
 - verzování (i metadat)
 - hromadné stažení (ZIP)

Děkuji za pozornost. Dotazy?

A decorative graphic at the bottom of the slide consists of several wavy lines. A prominent thick red wave is at the bottom, with a grey wave above it. Above these are several thin, light blue wavy lines. Three small blue dots are placed at various points along these lines, with thin lines connecting them to the main wave structure.

Vlastimil Krejčíř
krejcir@ics.muni.cz