



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

### OAI-PHM

Smyslem platformy otevřených archivů (OAI) je zpřístupnění na Webu dostupných materiálů prostřednictvím sdílení metadat mezi repozitáři, jejich publikování a archivace. Celá platforma přitom pochází z komunitní potřeby publikování vědeckých prací v elektronické podobě (e-prints). Cílem OAI je vytvořit výměnný rámec s příslušnými standardy, který by umožňoval efektivní šíření a sdílení obsahu napříč jednotlivými technickými řešeními.

OAI-PMH je protokol pro sklízení metadat, který definuje mechanismus sklízení záznamů z různých repozitářů, přičemž využívá robustní XML a HTTP specifikace. Přitom platí, že na straně serveru musí být podporován alespoň nekvalifikovaný Dublin Code. Samotný protokol přitom nspecifikuje vazbu mezi metadaty a obsahem ani nedává možnost vyhledávání. Tyto služby musí být poskytovány dalšími prostředky, které s OAI-PHM spolupracují. Samotný protokol jen sklízí metadata do jednoho místa. Ač je myšlenka sdílení a výměny metadat v zásadě velice jednoduchá a má celou řadu praktických výhod (například získávání skrytého obsahu), její praktické nasazení a implementace tak, aby celý systém smysluplně fungoval, je stále velice složitou záležitostí.

Neméně zajímavá je také historie vzniku tohoto formátu. Jak jsme již naznačili, vznikl z komunitní potřeby sklízet metadata, která byla umístěna v jednotlivých repozitářích, do kterých je vkládali samotní uživatelé. Zřejmě největším a nejnámějším projektem tohoto typu je arXiv (lze avšak zmínit také například CogPrints, NCSTRL či RePEc), který obsahoval preprinty fyzikálních a mateematických článků. Cílem bylo zpřístupnění objevů a novinek dříve, než jsou klasicky publikovány a recenzovány, tedy akcelerace vědy za cenu neověřenosti výsledků.

Konference v Santa Fe se proto pokusila rozhodnout a definovat klíčové parametry, které se budou pro sklízení dat používat. Předně bylo potřeba provést volbu mezi formáty metadat Dublin Core a MARC (uspěl první – respektive původně OAMS), transportním protokolem (tím se stal HTTP na úkor FTP), rozhodnout co budou zdrojové dokumenty pro získávání dat (v Santa Fe konvence to byly přímo e-prints dnes jsou to resource), a určit volbu formátu odpovědi – což bylo (a doposud je) XML.

Samotný model přitom rozlišuje dvě základní role. První jsou poskytovatelé dat, což jsou nejrůznější archivy a uložistiště. Poskytovatelé služeb užívají OAI rozhraní pro sklízení dat a jejich ukládání. Poskytovatelé služeb pak nad svojí databází výsledků staví další služby, jako je vyhledávání, zpřístupňování dat či filtrování výsledků. Mimo to mohou existovat také agregátory, které nabízejí jen sklizeň a ukládání metadat pro poskytovatele služeb.

Velice užitečnou vlastností jsou sady záznamů, které umožňují logické členění repozitářů. Sady se vytvářejí na základě obsahové či jiné příbuznosti dokumentů a mohou být stahovány pouze ony, místo velkých repozitářů. Příkladem setu může být například množina všech diplomových prací nebo hudebních záznamů.



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

### Příklad toho, jak může vypadat takový záznam:

```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2013-05-15T11:16:35Z</responseDate>
  <request verb="GetRecord" identifier="oai:kramerius.nkp.cz:p-ABA001/1" metadataPrefix="oai_dc">http://kramerius.nkp.cz/kramerius/oai</request>
- <GetRecord>
- <record>
  - <header>
    <identifier>oai:kramerius.nkp.cz:p-ABA001/1</identifier>
    <datestamp>2008-04-30T11:13:46Z</datestamp>
    <setSpec>type:periodical</setSpec>
  </header>
- <metadata>
  - <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>Filmový kurýr</dc:title>
    <dc:description>periodikum</dc:description>
    <dc:publisher>Josef Koza</dc:publisher>
    <dc:publisher>Zemský svaz kinematografů v Čechách</dc:publisher>
    <dc:publisher>Českomoravské filmové ústředí v Praze</dc:publisher>
    <dc:type>Digitized Periodical</dc:type>
    <dc:format>Pages represented as DjVu images</dc:format>
    <dc:identifier>1801-9501</dc:identifier>
    <dc:identifier>http://kramerius.nkp.cz/kramerius/handle/ABA001/1</dc:identifier>
    <dc:language>cze</dc:language>
  - <dc:coverage>
    1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944
  </dc:coverage>
  - <dc:rights>
    Volumes published later than in 1880 are digitally available only in the institutions that hold the original document.
  </dc:rights>
  </oai_dc:dc>
  </metadata>
</record>
</GetRecord>
</OAI-PMH>
```

Podrobněji viz HANOUSEK, Tomáš. OAI-PMH pro začátečníky. Dostupné na [http://www.nacr.cz/Z-files/moznosti\\_06.pdf](http://www.nacr.cz/Z-files/moznosti_06.pdf)