

OAI-PMH -haravointirajapinta Finnan indeksisiin

See [OAI-PMH Harvesting Interface for Finna's Index](#) for English version.

Historia

v1.0 Ensimmäinen versio

Yleistä

[OAI-PMH](#) on yleisesti käytetty, erityisesti metadatan haravointiin suunniteltu rajapinta. Finnassa OAI-PMH:lla voi haravoida Finnasta löytyvät vapaasti käytettävät materiaalit. OAI-PMH -haravointiin tarvitaan protokollaa tukeva ohjelma, ns. harvester. Valmiita kirjastoja on saatavilla eri ohjelmointikielille runsaasti, ja myös valmiita OAI-PMH:ta tukevia ohjelmistoja löytyy. Finnassa käytetään Kansalliskirjastossa kehitettyä [RecordManageria](#), ja openarchives.org:n sivuilta löytyy myös [useita vaihtoehtoja](#).

OAI-PMH:ssa olennaista on, että haravointi tapahtuu erissä, esim. 1000 tietuetta kerralla. Tällöin sekä haravoija että tarjoaja voivat käsitellä pyynnön ja vastauksen kohtuullisessa ajassa. Jokaisessa vastauksessa on mukana ns. resumptionToken, jolla saa pyydettyä seuraavan erän tietueita, kunnes enempää tietueita ei ole enää tarjolla. Suosittelemme käyttämään haravoijaa, joka osaa yrittää pyyntöä uudelleen hetken kuluttua, jos se katkeaa esim. verkkovirheeseen. Lisäksi on huomioitava, että yksittäinenkin vastaus voi olla melko suuri tietueiden sisällöstä riippuen, ja esim. libxml2-pohjaista parseria käyttäessä voi tarvita XML_PARSE_HUGE-optiota.

Finnan OAI-PMH -rajapinta löytyy osoitteesta <https://api.finna.fi/OAI/Server>.

Rajoitukset

- Haravoitava kokonaisuus voidaan rajata käyttämällä tarjoajan määrittelemiä settejä. OAI-PMH:ssa ei ole hakusanoja tai muita suodatustapoja.
- Finna ei tue poistettujen tietueiden raportointia.
- Haravoijan on raportoitava User-Agent, jota ei tulkita botiksi. Finna käyttää tähän [Crawler-Detect](#) -moduulia. Muuten vastauksena on status-koodi 403 (Forbidden). On suositeltavaa käyttää palvelukohtaista User-Agentia.

Finnan metadataformaattit

Finnassa on metadataa useissa eri formaateissa. OAI-PMH:lla on mahdollista haravoida kaikki tietyn lähtöformaatin tietueet, perustiedot Dublin Core -muodossa tai erityistä yhdistelmäformaattia, johon on mapattu Finnassa käsiteltyä metadataa. Taulukon kokoelmalistaus-linkit vievät Finnan [hakurajapinnar](#) puolelle.

Prefiksi	Formaatti	Kattavuus	Kuvaus
oai_dc	Dublin Core	Kaikki sisältö	Perusformaatti, jossa voi haravoida kaikki tietueet (kokoelmalistaus)
oai_vufind_json	Finnan oma	Kaikki sisältö	Yhdistelmä Dublin Core -perustietueesta ja lisätiedoista JSON-muodossa oai_vufind_json:metadata-elementin sisällä. Kuvaus tarjottavien kenttien sisällöstä löytyy Finnan hakurajapinnan dokumentaatiosta . Tarjolla on kuitenkin vain osa hakurajapinnan kentistä. Sisältö vaihtelee lähtöformaattista riippuen, eikä kaikissa tietueissa ole sisältöä kaikissa kentissä.
marc21	MARCXML	Lähinnä kirjastojen aineistoa (kirjastoluettelot)	Erityisesti kirjastoaineiston kuvailussa käytetty metadataformaatti (kokoelmalistaus)
oai_ead	EAD	Arkistojen aineistoa	Arkistoaineiston kuvailuun tarkoitettu formaatti, vanha versio (kokoelmalistaus)
oai_ead3	EAD3		Arkistoaineiston kuvailuun tarkoitettu formaatti, uusi versio (kokoelmalistaus)
oai_forward	FORWARD	Kansallisen audiovisuaalisen instituutin aineisto	EN15907 -standardiin pohjautuva esitysmuoto audiovisuaaliselle aineistolle (kokoelmalistaus)
oai_lido	LIDO	Museoiden aineistoa	Erityisesti museoaineiston kuvailussa käytetty metadataformaatti (kokoelmalistaus)
oai_qdc	Qualified Dublin Core	Julkaisuarkistojen aineistoa, opinnäytteitä, kirjastojen aineistoja	Dublin Corea laajempi muoto, jota käytetään usein mm. julkaisuarkistoissa (kokoelmalistaus)

Finnan setit

Pyytämällä tietueet ilman settimäärittystä saa edustavat tietueet Finnan deduplikoidusta kokonaisuudesta. Tarjolla on myös seuraavat setit:

Setti	Kuvaus
-------	--------

non_dedup	Deduplikoimaton kokonaisuus. Suurempi kuin perussetti, eli sisältää kaikki duplikaatit.
-----------	---