Digital Access To Scholarship At Harvard (DASH) Dataset

Description:

DASH is Harvard's digital repository for scholarly articles, theses and dissertatinos, and other Harvard-affiliate generated literature. Harvard Library makes the bibliographic data openly available for all uses, with a standard set of APIs.

License:

Information about licensed rights to the DASH dataset is provided in the DASH Terms of Use.

API Access:

DASH supports two standard APIs for extracting article information: opensearch and OAI-PMH.

OpenSearch API

As of the November 9, 2009 (1.1.8) release, DASH includes an OpenSearch interface. OpenSearch is a RESTful Web service that performs a query and returns search results as RSS or ATOM feeds. It provides complete, unmediated access to the full power of DSpace's Lucene search engine; the UI places inherent limitations on the queries you can construct.

OpenSearch URLs

The URL of an OpenSearch request starts with **http://dash.harvard.edu/open-search/** (the trailing "/" is required). All parameters are specified as query arguments after that:

Parameter Name	Description
query	Lucene query string (see below)
format	Output format, must be one of: **atom**, **rss**, or one of the supported specific format versions, e.g. **atom_1.0, rss_1.0, rss_2.0**. *No default.*
scope	Search is restricted to a collection or community with the indicated handle
rpp	Number indicating the number of results per page (i.e. per request). Default is **10**. Specifying **0** invokes the default, so to get *all* results use an improbably large rpp, e.g. 500.
start	Number of the page to start with (if paginating results)
sort_by	Index of sorting criteria (same as DSpace advanced search values). Must match a sort-option index in the DSpace configuration. Currently they are: 0 - by relevance (default) 1 - by title 2 - by date of issue 3 - by date of accession (i.e. submit date)

Datasets

Bibliographic Dataset
DASH

Media

Press Release
Frequently Asked Questions

order	Ordering of sorted entries, either **ascending** or **descending**. Only effective when sort_by is nonzero.

Lucene Query Syntax

The query string specifies what field (index) to match with a value. It also supports Boolean combinations, ranges, proximity, and other advanced features. The value of the **query** parameter must be in the Lucene query language (suitably escaped for inclusion in a URL, of course).

Indexes

DASH is configured with the following indexes. Not all of them appear on the Advanced Search page.

OpenSearch/Lucene name	**Advanced Search**	**Description**
default	Full Text	All metadata indexes *and* extracted text.
author	Author	Keyword in author name, wildcards and phrases acceptable.
author_authority	*n/a*	Authority key value of a Harvard-affiliated author.
title	Title	Keyword or phrase in title of article or journal.
subject	Keyword	Subject keywords (dc.subject.* fields)
abstract	Abstract	Keyword or phrase in the dc.description.abstract metadata field.
fasDepartment	FAS Department	Keyword or phrase matching FAS Department name
identifier	Identifier	Any of the identifiers such as DOIs and URLs associated with the work, including the published version, other sources, and the DASH NRS identifier.
issued	Issue Date	Date of issue (original publication), as full date.
issued.year	*n/a*	Year only of the date of issue.
accessioned	*n/a*	Date of accession, i.e. when the submission is entered into the archive.
accessioned.year	*n/a*	Year only of the date of accession.

Search Hints

- When matching FAS departments, search for an unambiguous phrase, e.g. **fasDepartment:"Molecular and Cellular Biology",** not just **fasDepartment:biology.**
- Specify search terms in lower case to match case-insensitively.
- Full dates are specified as YYYYMMDD, e.g. **20091031**. See the Lucene Query Language for further details about specifying ranges, etc.
- To match a Harvard author by authority key, go to the DASH UI and choose "Browse by Harvard Author" from the Options menu. Find your author, and copy the URL it links to. That URL will contain query arguments including a value for "**authority**", this is the authority key you have to look for. For example: http://dash.harvard.edu/browse?authority=**0aa57a41c53b1ffad57d51bb715886d7**&type=harvardAuthor

becomes:

http://dash.harvard.edu/open-search/?query=author_authority:**0aa57a41c53b1ffad57d51bb715886d7**

Examples

This gets an Atom feed for *all* of the articles by Stuart Shieber, using his Harvard-affiliated authority identification. They are ordered by descending issue date so newest papers appear first, although the year granularity may present a problem.

http://dash.harvard.edu/open-search/?query=author_authority:c9e989d522c0...

This shows all articles from the department of Molecular and Cellular Biology:

http://dash.harvard.edu/open-search/?query=fasDepartment:%22Molecular%20...

OAI-PMH API

We provide OAI-PMH access in conformity with its open standard.

As a DASH-specific example, here's an OAI-PMH url configured to show all articles in the Graduate School of Education collection ("set=hdl_1_3345928") for a specified date range:

http://dash.harvard.edu/oai/request?verb=ListRecords&metadataPrefix=dashrdf&set=hdl_1_3345928&from=2010-01-01&until=2011-01-01

The resulting XML file contains metadata about each article and a url for the dash abstract page.

Harvard Home Harvard Library Comments & Questions

© 2012 President and Fellows of Harvard College | Last modified April 10, 2012