

The OpenDOAR Application Programmers' Interface (API) – Version 1.3

Peter Millington
University of Nottingham
16th July 2007

Basic Principles

The API works by client programs making HTTP get requests to <http://opendoar.org/api.php> or <http://opendoar.org/api13.php>, with the URL's arguments indicating the required query criteria and output format. The results are returned as an XML stream. The URL `api13.php` is specific to version 1.3 of the API. The URL `api.php` always uses the version that is current at runtime.

Licencing

OpenDOAR API output data is licenced for re-use under a Creative Commons Attribution-Non-Commercial-Share Alike licence. See: <http://creativecommons.org/licenses/by-nc-sa/2.0/uk/>.

Input Parameters

All URL parameter names are in lower case. Parameter values are case-insentitive.

Query Specification

1. `all` (*optional*) – **Complete database**. Only permitted value is '`all=y`'.
2. `rid` (*optional*) – **OpenDOAR Record Identifier**. A comma-separated list of internal OpenDOAR ID numbers. (The `rID` argument in the `<repository>` XML tag).
3. `co` (*optional*) – **Country/Continent**. A list of ISO-3166 2-letter country codes or continent/Region names (see Appendix A) separated by commas. These will be searched using OR logic. See: <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>
4. `la` (*optional*) – **Language**. A list of ISO 2-letter language codes separated by commas. These will be searched using OR logic. See: http://en.wikipedia.org/wiki/List_of_ISO_639-1_codes
5. `subject` (*optional*) – **Subject area codes**. A list of OpenDOAR subject codes, separated by commas. Multiple codes are searched using OR logic. List in Appendix B.
6. `rt` (*optional*) – **Repository type codes**. A list of OpenDOAR numeric repository type codes separated by commas. These will be searched using OR logic. List in Appendix C.
7. `ct` (*optional*) – **Content type codes**. A list of OpenDOAR numeric content type codes separated by commas. These will be searched using OR logic. List in Appendix D.
8. `pograde` (*optional*) – **Policy grade codes**. A list of OpenDOAR numeric policy grade codes separated by commas. These will be searched using OR logic. List in Appendix E.
9. `since` (*optional*) – **Update date**. Inclusive date since which the record was added, updated or otherwise processed. Format: e.g. '2006-09-21'

10. **before** (*optional*) – **Update date**. Exclusive date before which the record was added, updated or otherwise processed. Format: e.g. ‘2003-01-31’
11. **oai** (*optional*) – **OAI Base URL Availability**.
 - **y** – OAI Base URL available in *OpenDOAR*
 - **n** – OAI Base URL absent from *OpenDOAR*
12. **kwld** (*optional*) – **Keywords**. A list of space-separated strings for searching one or more of the text fields – e.g. repository name, organisation name, software name, etc. Case insensitive.
 - Spaces and HTML special characters should be URL-encoded – e.g. `&kwld=texas%20a%26m` (for Texas A&M).
 - Accented Latin characters are acceptable – e.g. `&kwld=qu%C3%A9bec`.
 - Non-latin scripts should be encoded using UTF-8 – e.g.:
 - `&kwld=%E4%BA%AC%E9%83%BD%E5%A4%A7%E5%AD%A6`
(for 京都大学, Kyoto)
 - `&kwld=%CE%A8%CE%B7%CF%86%CE%AF%CE%B4%CE%B1`
(for Ψηφίδας, Psepheda).
13. **multi** (*optional*) - **Qualifier for multi-string kwld parameters**. This may have one of the following values:
 - **phrase** [*default*] – treats everything as a single string or phrase, which may appear anywhere in the relevant fields. Any spaces are significant.
 - **begins** – String must appear at the start of the relevant fields. Spaces are significant.
 - **ends** – String must appear at the end of the relevant fields. Spaces are significant.
 - **all** – All the words must appear in the relevant field(s) – in any order
 - **any** – One or more of the words must appear in the relevant field(s)
14. **fields** (*optional*) – **Field family code**. This indicates which group of fields is to be searched by the *kwld* parameter. Only one option may be specified:
 - **rname** – Repository name, acronym or URL.
 - **oname** – Organisation name, acronym or URL – and Unit/Dept name, acronym or URL.
 - **names** – Repository name, Organisation name, Unit/Dept name, or their acronyms. (N.b. *not* URLs)
 - **sw** – Software name
 - **url** – Repository, OAI Base, Organisation, and Unit/Dept URLs.
 - **all** [*default*] – All the above (except ‘url’), plus Description and Remarks
15. **bool** (*optional*) – **Boolean logical operator**. This may be used if two or more different query options have been specified:
 - **and** [*default*] – Both or all query items must be found.
 - **or** – At least one of the query items must be found.
 - **[not** – not to be implemented (yet).]

Output Options

16. *show* (optional) – Collections of fields to be output:

- **min** – Minimum data – Repository name, acronym, preference, URL, and OAI Base URL. This is the minimum required for a list of repository titles, or for OAI harvesting.
- **basic** [default] – as for **min**, plus Academic unit name, acronym, preference and URL, Organisation name, acronym, preference and URL, Postal Address, Country, Latitude and Longitude, Phone and Fax, Description, Remarks, Number of records and Date harvested for record count. These are generally the most popular fields.
- **index** – as for **min**, plus lists of index terms and categories - i.e. Year established, Repository type, Operational status, Software name & version, Subject classes, Languages, Content types, and Policy grades.
- **contact** – as for **min**, plus list of contacts – i.e. Personal name, Job title, Email address and Phone number.
- **policy** – as for **min**, plus list of Policy grades, and Standardised policy statements.
- **max** – Maximum data for all available fields.

Multiple options can be specified, separated by commas. Where options overlap, fields are not duplicated in the output. See example output in Appendix F

17. *sort* (optional) – Sort order. A list of field codes in the order required, separated by commas: **rname** [default], **co**, **oname**, or **sw** - i.e. Repository name, Country, Organisation name, or Software name.

18. *header* (optional) – XML header. Controls whether or not to display an XML header section echoing the query and providing statistics and error reports on the outcome:

- **y** – Display a header.
- **n** [default] – Suppress the header

Constraints

- At least one query option must be specified (points 1 to 11).
- Parameters cannot have null (empty) values.
- If either the 'multi' or 'fields' parameters is specified, the 'kwd' parameter must also be specified.
- 'multi' and 'bool' expect multiple words and/or query items. No error is reported if there is only one.

Outcome and Error handling

- If 'header=y' is specified, an XML header section will: echo the query, report the number of hits retrieved (which could be zero), give textual error reports if anything untoward happens
- If 'header=n' and either nothing is found, or an error occurs, an empty XML file will be returned. The application may therefore need to repeat the query with 'header=y' if the cause needs identifying.

Examples

- <http://opendoar.org/api13.php?co=fr.ca.be.ch&la=fr&show=min&sort=co.rname>
Finds repositories in France, Canada, Belgium or Switzerland in the French language. Returns repository names & URLs sorted by country & repository name.
- <http://opendoar.org/api13.php?kwd=geo%20earth&multi=any&subject=cam&bool=or>
Finds repositories containing the strings 'geo' or 'earth' or with the subject code for 'earth and planetary sciences'. By default, returns summary records sorted by repository name.
- <http://opendoar.org/api13.php?kwd=CCLRC%20ePublication%20Archive&show=max>
Shows all available OpenDOAR data for the *CCLRC ePublication Archive*.

Changes since Version 1.2

- The base URL has been changed from `api12.php` to `api13.php` to reflect the new version number.
- A new query option has been added - `rt` - that searches for Repository Type code.

Appendix A – OpenDOAR Continent/Region Names

Africa
 Antarctica *
 Asia
 Australasia
 Caribbean *
 Central America
 Europe
 North America
 Oceania *
 South America

* At the time of writing, there were no repositories in OpenDOAR for these continents.

Appendix B – Hierarchical List of OpenDOAR Subject Area Codes

C	Multidisciplinary
Ca	Science General
Cad	Agriculture, Food and Veterinary
Cag	Biology and Biochemistry
Cak	Chemistry and Chemical Technology
Cam	Earth and Planetary Sciences
Cap	Ecology and Environment
Car	Mathematics and Statistics
Cat	Physics and Astronomy
Ce	Health and Medicine
Ci	Technology General
Cif	Architecture
Cil	Civil Engineering
Cin	Computers and IT
Cis	Electrical and Electronic Engineering
Cix	Mechanical Engineering and Materials
Co	Arts and Humanities General
Cod	Fine and Performing Arts
Cog	Geography and Regional Studies
Cok	History and Archaeology
Col	Language and Literature
Cop	Philosophy and Religion
Cu	Social Sciences General
Cub	Business and Economics
Cud	Education
Cup	Law and Politics
Cur	Library and Information Science
Cut	Management and Planning
Cuv	Psychology

Appendix C – OpenDOAR Repository Type Codes

1	Undetermined	- Repositories whose type has not yet been assessed
2	Institutional	- Institutional or departmental repositories
3	Disciplinary	- Cross-institutional subject repositories
4	Aggregating	- Archives aggregating data from several subsidiary repositories
5	Governmental	- Repositories for governmental data

Appendix D – OpenDOAR Content Type Codes

1	Research papers (pre- and postprints)
2	Research papers (preprints only)
3	Research papers (postprints only)
4	Bibliographic references
5	Conference and workshop papers
6	Theses and dissertations
7	Unpublished reports and working papers
8	Books chapters and sections
9	Datasets
10	Learning Objects
11	Multimedia and audio-visual materials
12	Software
13	Patents
14	Other special item types

Appendix E – OpenDOAR Policy Grade Codes

1	Content policies unknown
2	Content policies not stated
3	Content policies explicitly undefined
4	Content policies unclearly stated
5	Content policies defined
6	Metadata policies unknown
7	Metadata policies not stated
8	Metadata re-use policy explicitly undefined
9	All metadata re-use prohibited
10	Metadata re-use policy unclearly stated
11	Metadata re-use permitted for not-for-profit purposes
12	Commercial metadata re-use permitted
13	Full data item policies unknown
14	Full data item policies not stated
15	Full data item policies explicitly undefined
16	All re-use of full data items prohibited
17	Harvesting full data items by robots prohibited
18	Policy for the re-use of full data items unclearly stated
19	Rights vary for the re-use of full data items
20	Re-use of full data items permitted for not-for-profit purposes
21	Commercial re-use of full data items permitted
22	Submission policies unknown
23	Submission policies not stated
24	Submission policies explicitly undefined
25	Submission policies unclearly stated
26	Submission policies defined
27	Preservation policies unknown
28	Preservation policies not stated
29	Preservation policies explicitly undefined
30	Preservation policies unclearly stated
31	Preservation policies defined

Appendix F – Example Output XML record

XML header & Legal Notices

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE OpenDOAR SYSTEM "api13.dtd">
<OpenDOAR apiVersion="1.3">
  <copyright>Copyright 2007, University of Nottingham</copyright>
  <licence>OpenDOAR data is available for re-use under a Creative Commons
    Attribution-Non-Commercial-Share Alike licence</licence>
```

API Query Parameters & Diagnostics – Displayed when &header=y

```
<header>
  <parameters>
    <parameter>
      <parametername>kwd</parametername>
      <parametervalue>xerxes</parametervalue>
    </parameter>
    <parameter>
      <parametername>show</parametername>
      <parametervalue>max</parametervalue>
    </parameter>
    <parameter>
      <parametername>header</parametername>
      <parametervalue>y</parametervalue>
    </parameter>
    <parameter>
      <parametername>multi</parametername>
      <parametervalue>phrase</parametervalue>
    </parameter>
    <parameter>
      <parametername>fields</parametername>
      <parametervalue>all</parametervalue>
    </parameter>
    <parameter>
      <parametername>bool</parametername>
      <parametervalue>AND</parametervalue>
    </parameter>
    <parameter>
      <parametername>sort</parametername>
      <parametervalue>rname</parametervalue>
    </parameter>
  </parameters>
  <numhits>1</numhits>
  <timestamp>16-Jul-2007:16:28:14</timestamp>
  <errors />
</header>
```

Main repository listing

```
<repositories>
```

Data output for all &show options – the only data output for &show=min

```
<repository rID="572">
  <rName>Undergraduate theses &amp; essays from Lund University</rName>
  <rAcronym>Xerxes</rAcronym>
  <rNamePreferred>Y</rNamePreferred>
  <rUrl>http://theses.lub.lu.se/undergrad/</rUrl>
  <rOaiBaseUrl />
```


Additional data output for &show=basic and &show=max options

```

<uName>Lund University Digital Library</uName>
<uAcronym />
<uNamePreferred>Y</uNamePreferred>
<uUrl />
<oName>Lunds universitet</oName>
<oAcronym />
<oNamePreferred>Y</oNamePreferred>
<oUrl>http://www.lu.se/</oUrl>
<postalAddress>Lund University Libraries</postalAddress>
<country>
  <cIsoCode>SE</cIsoCode>
  <cCountry>Sweden</cCountry>
</country>
<paLatitude />
<paLongitude />
<paPhone>046-2224005</paPhone>
<paFax />
<rDescription>This is an institutional repository for essays and undergraduate
  dissertations from the university. While material has been deposited by
  members from all departments there are notably more from those disciplines
  associated with the social sciences. The site interface is available in
  English or Swedish, and abstracts for materials are provided in both
  languages. An RSS feed is available for anyone interested in keeping
  up-to-date with newly added materials.&lt;br&gt;Xerxes is part of the national
  database for undergraduate essays, Uppsok, which is part of
  Libris.</rDescription>
<rRemarks>Partners: Uppsok. Special items include: Student essays</rRemarks>
<rNumOfItems>6292</rNumOfItems>
<rDateHarvested>2007-07-16</rDateHarvested>

```

Additional data output for &show=index and &show=max options

```

<rYearEstablished />
<repositoryType>Institutional</repositoryType>
<operationalStatus>Operational</operationalStatus>
<rSoftWareName />
<rSoftWareVersion />
<classes>
  <class>
    <clCode>C</clCode>
    <clTitle>Multidisciplinary</clTitle>
  </class>
</classes>
<languages>
  <language>
    <lIsoCode>en</lIsoCode>
    <lName>English</lName>
  </language>
  <language>
    <lIsoCode>sv</lIsoCode>
    <lName>Swedish</lName>
  </language>
</languages>
<contentTypes>
  <contentType ctID="6">Theses and dissertations</contentType>
  <contentType ctID="14">Other special item types</contentType>
</contentTypes>

```

policyType & policyGrade output for &show=index and &show=max options

```

<policies>
  <policy>
    <policyType potID="1">Content</policyType>
    <policyGrade pogID="5">Content policies defined</policyGrade>
  </policy>
</policies>

```

poStandard additionally output for &show=policy and &show=max options

```

<poStandard>
  <item>This is an institutional or departmental repository.</item>
  <item>The repository is restricted to: Theses and dissertations</item>
</poStandard>
</policy>

<policy>
  <policyType potID="2">Metadata</policyType>
  <policyGrade pogID="8">Metadata re-use policy explicitly
    undefined</policyGrade>
  <poStandard>
    <item>Anyone may access the metadata free of charge.</item>
    <item>No metadata re-use policy defined. Assume no rights at all have been
      granted.</item>
  </poStandard>
</policy>

<policy>
  <policyType potID="3">Data</policyType>
  <policyGrade pogID="15">Full data item policies explicitly
    undefined</policyGrade>
  <poStandard>
    <item>Anyone may access full items free of charge.</item>
    <item>No full-item re-use policy defined. Assume no rights at all have
      been granted.</item>
  </poStandard>
</policy>

<policy>
  <policyType potID="4">Submission</policyType>
  <policyGrade pogID="26">Submission policies defined</policyGrade>
  <poStandard>
    <item>Items may only be deposited by registered students of the
      institution</item>
    <item>No moderation policy defined. Assume nothing has been vetted.</item>
    <item>The validity and authenticity of the content of submissions is the
      sole responsibility of the depositor.</item>
    <item>No embargo policy defined.</item>
  </poStandard>
</policy>

<policy>
  <policyType potID="5">Preserve</policyType>
  <policyGrade pogID="28">Preservation policies not stated</policyGrade>
  <poStandard>
    <item>No preservation policy defined.</item>
  </poStandard>
</policy>
</policies>

```

Additional data output for `&show=contact` and `&show=max` options

```
<contacts>
  <person>
    <pName />
    <pJobTitle>Repository Administrator</pJobTitle>
    <pEmail>xerxes-admin@munin.lub.lu.se</pEmail>
    <pPhone />
  </person>
</contacts>
```

End tags

```
</repository>
</repositories>
</OpenDOAR>
```

Appendix G – Document Type Definition (DTD) for XML output

```

<?xml version='1.0' encoding='UTF-8' ?>
<!--
  Document Type Definition for OpenDOAR API 1.3 output
  Copyright 2007 SHERPA, University of Nottingham
-->
<!-- <OpenDOAR> is the root element -->
<!ELEMENT OpenDOAR          (copyright,licence,header?,repositories)>
<!ATTLIST OpenDOAR apiVersion CDATA #REQUIRED>
<!ELEMENT copyright         (#PCDATA)>
<!ELEMENT licence          (#PCDATA)>
<!ELEMENT header           (parameters+,numhits?,timestamp,errors*)>
<!ELEMENT parameters       (parameter*)>
<!ELEMENT parameter        (parametername?,parametervalue?)>
<!ELEMENT parametername    (#PCDATA)>
<!ELEMENT parametervalue   (#PCDATA)>
<!ELEMENT numhits          (#PCDATA)>
<!ELEMENT timestamp        (#PCDATA)>
<!ELEMENT errors           (error*)>
<!ELEMENT error            (#PCDATA)>
<!ELEMENT repositories     (repository*)>
<!ELEMENT repository       (rName?,rAcronym?,rNamePreferred?,rUrl,
                             rOaiBaseUrl?,
                             uName?,uAcronym?,uNamePreferred?,uUrl?,
                             oName?,oAcronym?,oNamePreferred?,oUrl?,
                             postalAddress?,country,paPhone?,paFax?,
                             rDescription?,rRemarks?,rNumOfItems?,rDateHarvested?,
                             rYearEstablished?,repositoryType?,operationalStatus?,
                             rSoftWareName?,rSoftWareVersion?,
                             classes*,languages*,contentTypes*,policies*,contacts*)>
<!ATTLIST repository rID    CDATA #REQUIRED>
<!ELEMENT rName           (#PCDATA)>
<!ELEMENT rAcronym        (#PCDATA)>
<!ELEMENT rNamePreferred  (#PCDATA)>
<!ELEMENT rUrl            (#PCDATA)>
<!ELEMENT rOaiBaseUrl     (#PCDATA)>
<!ELEMENT uName           (#PCDATA)>
<!ELEMENT uAcronym        (#PCDATA)>
<!ELEMENT uNamePreferred  (#PCDATA)>
<!ELEMENT uUrl            (#PCDATA)>
<!ELEMENT oName           (#PCDATA)>
<!ELEMENT oAcronym        (#PCDATA)>
<!ELEMENT oNamePreferred  (#PCDATA)>
<!ELEMENT oUrl            (#PCDATA)>
<!ELEMENT postalAddress   (#PCDATA)>
<!ELEMENT country         (cIsoCode,cCountry)>
<!ELEMENT cIsoCode        (#PCDATA)>
<!ELEMENT cCountry        (#PCDATA)>
<!ELEMENT paPhone         (#PCDATA)>
<!ELEMENT paFax           (#PCDATA)>
<!ELEMENT rDescription    (#PCDATA)>
<!ELEMENT rRemarks       (#PCDATA)>
<!ELEMENT rNumOfItems     (#PCDATA)>
<!ELEMENT rDateHarvested  (#PCDATA)>
<!ELEMENT rYearEstablished (#PCDATA)>
<!ELEMENT repositoryType  (#PCDATA)>
<!ELEMENT operationalStatus (#PCDATA)>
<!ELEMENT rSoftWareName   (#PCDATA)>
<!ELEMENT rSoftWareVersion (#PCDATA)>
<!ELEMENT classes         (class*)>
<!ELEMENT class           (clCode,clTitle)>

```

```
<!ELEMENT clCode (#PCDATA)>
<!ELEMENT clTitle (#PCDATA)>
<!ELEMENT languages (language*)>
<!ELEMENT language (lIsoCode,lName)>
<!ELEMENT lIsoCode (#PCDATA)>
<!ELEMENT lName (#PCDATA)>
<!ELEMENT contentTypes (contentType*)>
<!ELEMENT contentType (#PCDATA)>
<!ATTLIST contentType ctID CDATA #REQUIRED>
<!ELEMENT policies (policy*)>
<!ELEMENT policy (policyType,policyGrade,poStandard?)>
<!ELEMENT policyType (#PCDATA)>
<!ATTLIST policyType potID CDATA #REQUIRED>
<!ELEMENT policyGrade (#PCDATA)>
<!ATTLIST policyGrade pogID CDATA #REQUIRED>
<!ELEMENT poStandard (item+)>
<!ELEMENT item (#PCDATA|a|subitems)*>
<!ELEMENT a (#PCDATA)>
<!ATTLIST a href CDATA #REQUIRED>
<!ELEMENT subitems (subitem+)>
<!ELEMENT subitem (#PCDATA)>
<!ELEMENT contacts (person*)>
<!ELEMENT person (pName?,pJobTitle?,pEmail?,pPhone?)>
<!ELEMENT pName (#PCDATA)>
<!ELEMENT pJobTitle (#PCDATA)>
<!ELEMENT pEmail (#PCDATA)>
<!ELEMENT pPhone (#PCDATA)>
```