

The SHERPA/RoMEO Application Programmers' Interface (API) Prototype Version 1.0

Circulation: PUBLIC
Peter Millington
University of Nottingham
21st April 2006

Introduction

The SHERPA/RoMEO Application Programmers' Interface (API) is a machine-to-machine interface that lets programmers access SHERPA/RoMEO data from their applications. For instance, they could use the API to incorporate an automatic look-up of a journal or publisher into a repository's deposition process.

Principles

Applications submit queries to SHERPA/RoMEO as HTTP requests, using the API's URL with appropriate query parameters attached.

e.g. <http://www.sherpa.ac.uk/romeoapi.php?pub=university%20press&qtype=all>

where:

- <http://www.sherpa.ac.uk/romeoapi.php> is the URL of the API
- **?pub=university%20press** is the first parameter. In this example, 'pub' is the name of the parameter (indicating a search for publisher's name), and 'university%20press' is its value.

Note that the parameters are URL-encoded, so that a space character is represented by the code '%20'.

- **&qtype=all** is the second parameter. 'qtype' is the name of the query type parameter, and in this case, 'all' indicates that all the strings specified in 'pub' must be present in the publisher's name

Note that the first URL parameter is signified by a '?' character. Subsequent parameters start instead with an '&'. The parameters can appear in any order, provided this convention is maintained. Thus '...?qtype=all&pub=university%20press' is also valid.

The HTML requests are sent to the SHERPA server where they are run against the SHERPA/RoMEO database, and the results returned as an XML file.

For test purposes, API requests may be entered into the address bar of a browser such as Internet Explorer or Firefox. Properly structured XML should be displayed

Example 1 is a sample of the API's output.

The XML document type definition is in file <http://www.sherpa.ac.uk/romeoapi.dtd> - see *Appendix A*

Four types of request can be made:

- Publisher name – Search using keywords and strings
- SHERPA/RoMEO Database Record ID – Search using ID number for the record for a single publisher
- Journal Title - Search using keywords and strings get a list of matching titles
- Journal Follow-up – Follow up a journal title search to get the publisher's information.

Example 1 - XML Output from the SHERPA/RoMEO API

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>jtitle</name>
        <value>journal of geology</value>
      </parameter>
    </parameters>
    <numhits>1</numhits>
    <apicontrol>journal</apicontrol>
    <outcome>singleJournal</outcome>
    <message />
    <timestamp>15-Apr-2006:09:19:48</timestamp>
  </header>
  <journals>
    <journal>
      <jtitle>Journal of geology -chicago-</jtitle>
      <issn>0022-1376</issn>
      <zetocpub>UNIVERSITY OF CHICAGO PRESS</zetocpub>
      <romeopub>University of Chicago Press</romeopub>
    </journal>
  </journals>
  <publishers>
    <publisher id="77">
      <name>University of Chicago Press</name>
      <alias />
      <homeurl>http://www.press.uchicago.edu/</homeurl>
      <preprints>
        <archiving>can</archiving>
        <restrictions />
      </preprints>
      <postprints>
        <archiving>conditional</archiving>
        <restrictions>
          <restriction>Some journals impose embargoes</restriction>
          <restriction>Consult journal home page for details</restriction>
        </restrictions>
      </postprints>
      <conditions>
        <condition>On a non-profit author, institutional or discipline-
          specific pre-print server</condition>
        <condition>Publisher copyright and source must be
          acknowledged</condition>
        <condition>Authors are encouraged to use publisher version (PDF) or
          link to publisher version</condition>
      </conditions>
      <copyright>View &A
        HREF="http://www.journals.uchicago.edu/rights.html"&A;archiving
        policy&A; and &A
        HREF="http://www.journals.uchicago.edu/GenPubAgree.pdf"&A;sample
        publication agreement&A;</copyright>
      <romeocolour>yellow</romeocolour>
    </publisher>
  </publishers>
</romeoapi>

```

Explanation of the Main XML Elements

<parameters>

Each parameter consists of a <name>-<value> pair. Errors may occur if either of these is mis-spelled or if the <value> is empty.

<numhits>

If the query is valid, this returns the number of relevant records found – which may of course be '0'. If the query is invalid, or if an error occurs, numhits is returned empty.

<apicontrol> and <outcome>

These elements are generated by the API while processing the query.

<apicontrol> may have the value: *followup*, *identifier*, *invalid*, *journal*, or *publisher*.

<outcome> may have the value: *failed*, *manyJournals*, *notFound*, *publisherFound*, *singleJournal*, or *uniqueZetoc*

These elements are unlikely to be of use to application developers, except for debugging.

<message>

This element holds any error and warning messages.

<journals>

Each <journal> record always consists of four elements

- <jtitle> - Journal title
- <issn> - International Standard Serial Number *
- <zetocpub> - Publisher's name, as found in ZETOC *
- <sherpapub> - Publisher's name in the RoMEO database *

Asterisked fields may be empty if either the journal is not found in ZETOC and/or the publisher is not in the SHERPA database.

A non-null <sherpapub> element indicates that the publisher is covered by the SHERPA/RoMEO database, and can therefore be used for a follow-up publisher search. The other three elements derive from the ZETOC database. The ZETOC and RoMEO publishers may differ. Publishers' names in RoMEO are standardised and may also cover subsidiaries.

<preprints>

<archiving> indicates whether or not authors may self-archive pre-prints (subject to any conditions given later in the record). It may have one of four values: *can*, *cannot*, *restricted*, or *unclear*. In the manual SHERPA/RoMEO system, these correspond with the symbols: green tick, red cross, dashed grey tick, and question mark, respectively.

If self-archiving is *restricted*, the <restrictions> element will contain a list of the publisher's restrictions, otherwise, this element is left empty.

<postprints>

The pair of elements in this section is the same as for pre-prints.

<conditions>

These may apply to pre-prints and/or post-prints

<copyright>

One or more hyperlinks to the publisher's copyright policy statements - if available online.

Publisher Name:**Parameters:**

- **pub** – Required: query string(s) for a publisher search
- **qtype** – Optional: permitted values 'all' (default), 'any', or 'exact'

Example requests:

The following examples explain and illustrate the effects of the three possible `qtype` parameters:

- **qtype=all** – All the strings in `pub` must appear in the publisher's name, but they may be in any order or location. This is the default `qtype`.

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?pub=optical%20society&qtype=all>

Finds:

International Society for Optical Engineering

Optical Society of America

- **qtype=any** – Publishers' names must include one or more of the `pub` strings

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?pub=chemistry%20society&qtype=any>

Results include:

Biophysical Society

International Union of Pure and Applied Chemistry

Royal Society of Chemistry

- **qtype=exact** – The `pub` string must appear as an intact 'phrase' somewhere in the publisher's name.

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?pub=tute%20of%20physics&qtype=exact>

Only finds:

American Institute of Physics

Institute of Physics

Results:

Requests may retrieve zero, one or more results. If a single publisher is found, you should still check that it is the correct one. *Example 2* illustrates the results from the publisher search.

Note that an empty `<journals>` section is always returned with publisher searches.

Example 2 – Sample output from an API Publisher Search**Results:**

```
<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>pub</name>
        <value>institute of physics</value>
      </parameter>
    </parameters>
```

```

<numhits>2</numhits>
<apicontrol>publisher</apicontrol>
<outcome>publisherFound</outcome>
<message>Searching for publisher: 'institute+of+physics', query type:
  'ALL'</message>
<timestamp>18-Apr-2006:12:00:35</timestamp>
</header>
<journals />
<publishers>
  <publisher id="7">
    <name>American Institute of Physics</name>
    <alias />
    <homeurl>http://www.aip.org/</homeurl>
    <preprints>
      <archiving>can</archiving>
      <restrictions />
    </preprints>
    <postprints>
      <archiving>can</archiving>
      <restrictions />
    </postprints>
    <conditions>
      <condition>Publisher's version/PDF may be used on author's personal
        website only</condition>
      <condition>Author's own version of final article on e-print
        servers</condition>
      <condition>Must link to publisher version or journal home
        page</condition>
    </conditions>
    <copyright>ftp://ftp.aip.org/aipdocs/forms/copyright.pdf</copyright>
    <romeocolour>green</romeocolour>
  </publisher>
  <publisher id="40">
    <name>Institute of Physics</name>
    <alias />
    <homeurl>http://www.iop.org/</homeurl>
    <preprints>
      <archiving>can</archiving>
      <restrictions />
    </preprints>
    <postprints>
      <archiving>can</archiving>
      <restrictions />
    </postprints>
    <conditions>
      <condition>Published source must be acknowledged</condition>
      <condition>Publisher's version/PDF cannot be used</condition>
      <condition>Eprint server is non-profit</condition>
    </conditions>
    <copyright>View &A HREF="http://authors.iop.org/atom/help.nsf/0/
      F20EC7D4A1A670AA80256F1C0053EEFF?OpenDocument"&gt;
      policy&A&gt;</copyright>
    <romeocolour>green</romeocolour>
  </publisher>
</publishers>
</romeoapi>

```

SHERPA/RoMEO Database Record ID

Parameter:

- **id** – Required: Identifier of the required SHERPA/RoMEO database record

In *Example 2*, you can see that the <publisher> tag includes an **id** attribute, which is the publisher's internal record number for the SHERPA/RoMEO database. This can be used for a follow-up search to retrieve the single record, as follows:

e.g. <http://www.sherpa.ac.uk/romeoapi.php?id=3>

Example 3 – Sample Output from an API ID Search

```
<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>id</name>
        <value>3</value>
      </parameter>
    </parameters>
    <numhits>1</numhits>
    <apicontrol>identifier</apicontrol>
    <outcome>publisherFound</outcome>
    <message />
    <timestamp>18-Apr-2006:14:29:32</timestamp>
  </header>
  <journals />
  <publishers>
    <publisher id="3">
      <name>American Association for the Advancement of Science</name>
      <alias />
      <homeurl>http://www.aaas.org</homeurl>
      <preprints>
        <archiving>cannot</archiving>
        <restrictions />
      </preprints>
      <postprints>
        <archiving>restricted</archiving>
        <restrictions>
          <restriction>May be posted at the time of publication in Science,
            but not sooner</restriction>
        </restrictions>
      </postprints>
      <conditions>
        <condition>On author's personal web site</condition>
        <condition>Author's own version of accepted paper</condition>
        <condition>Must link to publisher version</condition>
        <condition>Published source must be acknowledged</condition>
      </conditions>
      <copyright>http://www.sciencemag.org/feature/contribinfo/prep/
        lic_info.pdf; and
        http://www.sciencemag.org/feature/contribinfo/prep/license
        .pdf</copyright>
      <romeocolour>blue</romeocolour>
    </publisher>
  </publishers>
</romeoapi>
```

Warning: We do not guarantee that **ids** will be permanently linked to particular publishers.

Journal Titles:**Parameters:**

- **jtitle** – Required: query string(s) for a journal title search
- **qtype** – Optional: permitted values 'starts' (default), 'contains', or 'exact'

Example requests:

The following examples explain and illustrate the effects of the three possible `qtype` parameters.

N.b. for journals, multiple words in `qtype` are treated as a single string -:

- **qtype=starts** – The `jtitle` string must appear at the start of the journal title. This is the default `qtype`.

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?jtitle=folk&qtype=starts>

Results include:

Folk -kobenhavn-
Folklore
Folk music journal

- **qtype=contains** – The `jtitle` string may appear anywhere in the journal title

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?jtitle=folk&qtype=contains>

Results include:

Folk life
Norfolk archaeology
Southern folklore

- **qtype=exact** – Full journal title (excluding any added place of publication) must equal `jtitle`

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?jtitle=folk&qtype=exact>

Only finds:

Folk -kobenhavn-

Results

There are two formats for the results, depending on whether multiple journals or a single journal are found:

Example 4 shows the output for multiple journals.

Please note that an empty `<publishers>` section is always returned for multiple journal output.

Warning: Requests may retrieve up to 50 journal titles only. There is no facility to get further matches. If your request retrieves 50 journals, and your title is not among the results, you should re-formulate your query with additional title keywords and/or a different `qtype` so that a smaller set is retrieved:

e.g.: <http://www.sherpa.ac.uk/romeoapi.php?jtitle=modern%20language&qtype=contains>

finds the journal 'Canadian modern language review', but

<http://www.sherpa.ac.uk/romeoapi.php?jtitle=modern&qtype=contains> does not, because it does not appear in the first 50 hits.

Example 4 – Sample Output from the API for Multiple Journals

```

<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>jtitle</name>
        <value>dna</value>
      </parameter>
    </parameters>
    <numhits>4</numhits>
    <apicontrol>journal</apicontrol>
    <outcome>manyJournals</outcome>
    <message />
    <timestamp>18-Apr-2006:11:40:28</timestamp>
  </header>
  <journals>
    <journal>
      <jtitle>Dna and cell biology</jtitle>
      <issn>1044-5498</issn>
      <zetocpub>Mary Ann Liebert%2C Inc.</zetocpub>
      <romeopub>Mary Ann Liebert</romeopub>
    </journal>
    <journal>
      <jtitle>Dna repair -amsterdam-</jtitle>
      <issn>1568-7864</issn>
      <zetocpub>Elsevier Science B.V.%2C Amsterdam.</zetocpub>
      <romeopub>Elsevier</romeopub>
    </journal>
    <journal>
      <jtitle>Dna research</jtitle>
      <issn>1340-2838</issn>
      <zetocpub>KAZUSA DNA RESEARCH INSTITUTE</zetocpub>
      <romeopub />
    </journal>
    <journal>
      <jtitle>Dna sequence</jtitle>
      <issn>1042-5179</issn>
      <zetocpub>TAYLOR & FRANCIS</zetocpub>
      <romeopub>Taylor & Francis</romeopub>
    </journal>
  </journals>
  <publishers />
</romeoapi>

```

If the `<romeopub>` field for a journal is not empty, the publisher is covered by the SHERPA/RoMEO database, and the data can therefore be used to specify a follow-up search for the publisher (see below).

The output is different if a journal title search finds a single journal. This is the layout shown in *Example 1*, which has a single record in the `<journals>` section, plus a single record in the `<publishers>` section.

Warning: If your request yields a single result, you or your end-user should still check that the journal title correctly matches the one you seek.

Journal Follow-up Query:

Provided the <romeopub> field is not empty, the values from a journal record can be used to get the publisher's policy details. This is equivalent to selecting one of the hits from a manual SHERPA/RoMEO journal search.

Parameters – which should be copied *literatim* from `jttitle` search results:

- **jttitle** – Required: the relevant journal title
- **zetocpub** – Required: Publisher's name from the ZETOC database
- **romeopub** – Required: Publisher's name from the SHERPA/RoMEO database
- **issn** – Optional: International Standard Serial Number

Results:

The output of a follow-up search is similar to the layout shown in *Example 1* – i.e. a single record in the <journals> section, plus a single record in the <publishers> section.

Unsuccessful Requests

If the required journal title or publisher is not present in the database, output similar to *Example 5* will be returned

Example 5 – Sample Output from the API for an Unsuccessful Request

```
<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>jttitle</name>
        <value>recycling journal</value>
      </parameter>
    </parameters>
    <numhits>0</numhits>
    <apicontrol>journal</apicontrol>
    <outcome>notFound</outcome>
    <message />
    <timestamp>18-Apr-2006:11:38:08</timestamp>
  </header>
  <journals />
  <publishers />
</romeoapi>
```

The key things to note for unsuccessful requests are:

- <numhits> is set to zero
- <outcome> is set to notFound.
- Both the <journals> and <publishers> sections are empty.

Invalid Requests

When a request is made to the API, the parameters are checked for certain simple errors. If an error is detected, the query is not run and the output returns diagnostic information, as in *Examples 6 & 7*. The important things to note are:

- `<numhits>` is empty – whereas a valid unsuccessful query would return zero.
- `<apicontrol>` is set to `invalid`
- `<outcome>` is set to `failed`
- `<message>` describes the error that has been detected.
- The `<journals>` and `<publishers>` sections are empty.

Example 6 – Sample Output from the API for an Invalid Request – No Parameters

```
<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
    </parameters>
    <numhits />
    <apicontrol>invalid</apicontrol>
    <outcome>failed</outcome>
    <message>Failed. No query parameters provided. Nothing entered, nothing
      gained.</message>
    <timestamp>19-Apr-2006:15:01:30</timestamp>
  </header>
  <journals />
  <publishers />
</romeoapi>
```

Example 7 – Sample Output from the API for an Invalid Request – Wrong Parameter

```
<!DOCTYPE note SYSTEM "romeoapi.dtd"><?xml version="1.0"
  encoding="ISO-8859-1" ?>
<romeoapi version="1.0">
  <header>
    <parameters>
      <parameter>
        <name>jttitle</name>
        <value>zoos zoology</value>
      </parameter>
      <parameter>
        <name>qtype</name>
        <value>any</value>
      </parameter>
    </parameters>
    <numhits />
    <apicontrol>invalid</apicontrol>
    <outcome>failed</outcome>
    <message>Query type 'ANY' unrecognised for journal title
      searches</message>
    <timestamp>18-Apr-2006:13:52:50</timestamp>
  </header>
  <journals />
  <publishers />
</romeoapi>
```

We do not claim the error checks to be comprehensive. You may also get unexpected results if you fail to give a value for a parameter.

Conditions of Use

The information that we provide through the SHERPA/RoMEO API is given freely to interested parties for their re-use, although the following conditions apply:

- The information provided by the SHERPA/RoMEO service is not for commercial re-use.
- It would be appreciated if the following logo could be incorporated somewhere on a relevant public page which uses SHERPA/RoMEO information and made an active link to the following page: <http://www.sherpa.ac.uk/romeo.php>



There are several different sizes of logo available: wherever possible, please include the logo on the page where the information is presented.

- The SHERPA/RoMEO logo is available in various sizes, and can be downloaded from: <http://www.sherpa.ac.uk/images/romeologos.html>

Wherever possible, please include the logo on the page where the information is presented.

- If the SHERPA/RoMEO information is re-interpreted or modified, then we ask for the following text to be used in addition to the logo in an appropriate place:

"This information is derived from the SHERPA/RoMEO database and has been modified for use here."

Technical Constraints and Future Versions

Version 1.0 of the API is a prototype, built on the back of the manual SHERPA/RoMEO web application. It therefore suffers the same constraints. We may be able to remove these constraints and/or add further features in a future release of the API. Subject to demand and the agreement of our partners, these may include:

- Removal of the 50-title Limit for Journal Searches
- Addition of an ISSN query option
- A feedback facility to let end-users submit suggestions and corrections.

Feedback

We seek opinions and feedback on this prototype. We would also be grateful if developers could test it in their own environments and report the results to us. Please send feedback to Peter Millington (peter.millington@nottingham.ac.uk), tel. +44 (0) 115 846 8481.

Appendix A – romeoapi.dtd

```

<?xml version="1.0"?>
<!ELEMENT romeoapi      (#PCDATA)>
<!ELEMENT header       (parameters,journals,publishers)>
<!ELEMENT parameters   (parameter*,numhits,apicontrol,outcome,message*,
                        timestamp)>
<!ELEMENT parameter    (name?,value?)>
<!ELEMENT name         (#PCDATA)>
<!ELEMENT value        (#PCDATA)>
<!ELEMENT numhits     (#PCDATA)>
<!ELEMENT apicontrol  (#PCDATA)>
<!ELEMENT outcome     (#PCDATA)>
<!ELEMENT message     (#PCDATA)>
<!ELEMENT timestamp   (#PCDATA)>
<!ELEMENT journals    (journal+)>
<!ELEMENT journal     (jtitle,issn?,zetocpub?,romeopub?)>
<!ELEMENT jtitle      (#PCDATA)>
<!ELEMENT issn        (#PCDATA)>
<!ELEMENT zetocpub    (#PCDATA)>
<!ELEMENT romeopub    (#PCDATA)>
<!ELEMENT publishers  (publisher+)>
<!ELEMENT publisher   (name,alias?,homeurl,preprints,postprints,
                        conditions*,copyright,romeocolour)>
<!ELEMENT name        (#PCDATA)>
<!ELEMENT alias       (#PCDATA)>
<!ELEMENT homeurl     (#PCDATA)>
<!ELEMENT preprints   (archiving,restrictions*)>
<!ELEMENT archiving   (#PCDATA)>
<!ELEMENT restrictions (#PCDATA)>
<!ELEMENT postprints  (archiving,restrictions*)>
<!ELEMENT archiving   (#PCDATA)>
<!ELEMENT restrictions (#PCDATA)>
<!ELEMENT conditions  (#PCDATA)>
<!ELEMENT copyright   (#PCDATA)>
<!ELEMENT romeocolour (#PCDATA)>
<!ATTLIST romeoapi version CDATA #REQUIRED>
<!ATTLIST publisher id ID #REQUIRED>

```

SHERPA Document

The SHERPA/RoMEO Application Programmers' Interface (API)

Prototype Version 1.0

Peter Millington

SHERPA Technical Development Officer

University of Nottingham

21st April 2006