*67 Bricks API Documentation for Springer Nature*

Assignment:

# *ISRCTN*

Points of contact for questions

**Daniel Rendall**

| | |
|---|---|
| Client: | Springer Nature |
| Prepared By: | Daniel Rendall |
| Document Version: | 0.6 |
| Date: | 18/08/2023 |
| Document Status: | Draft |
| Prepared for: | |

# 1 Contents


1 Contents ..... 2
2 Document information ..... 4
2.1 Purpose ..... 4
2.2 Scope ..... 4
2.3 Definitions, Acronyms and Abbreviations ..... 4
2.4 References ..... 4
2.5 API History ..... 4
3 The API ..... 6
3.1 Retrieving a single trial ..... 6
3.2 Querying for multiple trials ..... 7
3.2.1 Constraints ..... 7
3.2.1.1 trialStatus ..... 8
3.2.1.2 conditionCategory ..... 8
3.2.1.3 condition ..... 8
3.2.1.4 recruitmentCountry ..... 8
3.2.1.5 ageRange ..... 9
3.2.1.6 intervention ..... 9
3.2.1.7 funderName ..... 9
3.2.1.8 outcomeMeasures ..... 9
3.2.1.9 inclusion ..... 10
3.2.1.10 exclusion ..... 10
3.2.1.11 gender ..... 10
3.2.1.12 phase ..... 10
3.2.1.13 recruitmentStatus ..... 10
3.2.1.14 overallStartDate ..... 10
3.2.1.15 overallEndDate ..... 10
3.2.1.16 dateApplied ..... 10
3.2.1.17 lastEdited ..... 11
3.2.1.18 sponsorOrganisation ..... 11
3.2.1.19 fundRef ..... 11
3.2.1.20 gridId ..... 11
3.2.1.21 fundingType ..... 11

3.2.1.22 title ..... 11
3.2.1.23 acronym ..... 11
3.2.2 Querying examples ..... 12
4 Formats ..... 14
4.1 Default ..... 14
4.2 UKCTG ..... 14
4.3 WHO ..... 14
4.4 Internal (deprecated) ..... 14

# 2 Document information

## 2.1 Purpose

This document describes the use of the API made available on the ISRCTN website to query and retrieve trial data.

It is intended for third parties who wish to make use of ISRCTN trial data.

## 2.2 Scope

This document describes the API and the data formats that it will return. It does not describe the internal representation of the trial data.

For the initial release of this documentation we have chosen not to make any changes to the API, although some areas for improvement have been identified. A future version of the API may address these.

## 2.3 Definitions, Acronyms and Abbreviations

BPOR – Be Part of Research, a website run by NIHR

GRID – Global Research Identifier Database (https://www.grid.ac/)

ISRCTN – International Standard Randomised Controlled Trial Number (as well as being the name of the system, all trials are identified uniquely by an ISRCTN)

NIHR – National Institute for Health Research

RTS – Reference and Terminology Service (https://www.nihr.ac.uk/documents/integrated-research-intelligence-system)

UKCTG – United Kingdom Clinical Trials Gateway, the old name for what is now BPOR

WHO – World Health Organization

## 2.4 References

This document expands on the description of the API that was given in the Admin User Guide (which was not intended for distribution to third parties).

The searching is implemented using the Search API in MarkLogic; the documentation for this may be useful: https://docs.marklogic.com/guide/search-dev/string-query

XML Schema date / time data types: https://www.w3schools.com/xml/schema_dtypes_date.asp

The application uses funder data downloaded from CrossRef:
http://data.crossref.org/fundingdata/registry

## 2.5 API History

| Application date | Notes |
|---|---|

| | |
|---|---|
| **3rd February 2021** | First complete documentation of API which has remained essentially unchanged since the site was first launched. |
| **10th May 2022** | API formats updated to include the new “default” |

# 3 The API

The ISRCTN API allows trial data to be queried. There are two calls, one allowing data for a specific trial (identified by ISRCTN) to be retrieved, the other allowing queries over the available trials.

Trial data is returned as XML, there are 4 formats to choose from: “default” (essentially the data as stored in ISRCTN minus irrelevant information and possibly with redacted personally identifiable information), “who”, “ukctg” and “internal”. The “internal” format is a legacy format used by one particular consumer and includes fields which no longer exist in the stored record but which have been recreated from fields which superseded them – we do not recommend downloading data in this format.

The API is publicly available; it is not subject to e.g. the requirement to have an API key. We would ask anyone using the API to be considerate of other site users and to heed the following guidelines:

- If your queries may return large numbers of trials, please consider splitting them into smaller queries by restricting them based on date ranges and running a number of smaller queries rather than one big one.
- If you are aiming to download data from the site as part of a regular batch process running overnight, please use a single-threaded approach that runs queries sequentially rather than attempting to parallelize them. Also, please try to pick a time that is not “obvious” (i.e. something like 01:17 rather than 01:00) to minimise the risk of lots of external systems hitting the site simultaneously.
- When experimenting with queries, either set the “limit” parameter to a low number or omit it to get the default of 10 results. The results document will tell how roughly how many documents matched the query.

The base URL for all API calls is the URL of the site (https://www.isrctn.com/)

Calls to the API currently return a response of 200 when successful. Any other status code (e.g. 400, 500) should be taken to indicate failure. This may be tightened up in a future API version.

## *3.1 Retrieving a single trial*

Endpoint: `/api/trial/<isrctn>/format/<format>`

The `<isrctn>` is the ISRCTN for the trial; this will be a number e.g. “12345678” or a string with an ISRCTN prefix e.g. “ISRCTN12345678”. These are completely equivalent; if the ISRCTN prefix is present it will be stripped off internally.

The `<format>` must be one of the three strings {“internal”, “ukctg”, “who”} (for more information about formats see section 0 below.

Examples:

`/api/trial/17723526/format/internal`

⇨ Retrieves the trial with ISRCTN 17723526 in the internal format.

`/api/trial/ISRCTN49306926/format/who`

- ⇨ Retrieves the trial with ISRCTN 49306926 in the WHO format

## 3.2 Querying for multiple trials

Endpoint: `/api/query/format/<format>?q=<query>&limit=<limit>`

The `<format>` must be one of the three strings {“internal”, “ukctg”, “who”} (for more information about formats see section 0 below.

The `<limit>` must be an integer and represents the maximum number of results to be returned, defaulting to 10

The `<query>` is the query string to be used. This is passed, with some modifications and sanitisation to the underlying MarkLogic Search API.

A query consists of keywords and constraints, linked by Boolean operators such as AND and OR. A constraint can test for a particular value in a field, or for a value that is less than / greater than a particular value; the behaviour depends on the constraint and is documented below. Keywords unqualified by a constraint are searched for across the various text fields in the trial.

Examples of constraint usage:

recruitmentCountry: "United Kingdom"

- ⇨ The field “recruitmentCountry” must be equal to the value “United Kingdom”; note that if the value contains spaces it must be quoted.

lastEdited GE 2021-01-20T00:00:00

- ⇨ The field “lastEdited” must have a value that is greater than or equal to “2021-01-20T00:00:00”. Note that the underlying value is stored as an xs:dateTime with the time part set to 00:00:00 – you must include the time part in the query.

condition: dementia

- ⇨ The word “dementia” must appear in the free-text description of the condition covered by the trial.

### 3.2.1 Constraints

If the field type is “keyword”, the field matches an extended piece of text, such as the trialist’s description of the condition covered by the trial. It makes sense to do keyword searches on such fields. In general these are performed case-insensitively.

If the field type is “range”, the field matches a single specific datum in the trial – for example a metadata value taken from a controlled vocabulary, or a date such as the recruitment start date. If the data type for such a field is “string” then it makes sense to do exact value comparisons (e.g. that the recruitmentCountry is “United Kingdom”). If it is “dateTime” then it makes sense to do less than / greater than comparisons.

The operators which can be used in comparisons are LT, LE, GT, GE and NE (less than, less than or equal, greater than, greater than or equal, and not equal)

#### 3.2.1.1 trialStatus

Type: range(string)

For most trials, this is either “Ongoing” or “Completed”, which is computed based on whether the trial end date has been reached. However, it can be overridden by an editorially set value; currently these values include “Stopped”, “Suspended” and “Enrolling by invitation”

#### 3.2.1.2 conditionCategory

Type: range(string)

Category for the condition covered in the trial. It will be one of the following values:

"Not Specified", "Cancer", "Circulatory System", "Digestive System", "Ear, Nose and Throat", "Eye Diseases", "Genetic Diseases", "Haematological Disorders", "Infections and Infestations", "Injury, Occupational Diseases, Poisoning", "Mental and Behavioural Disorders", "Musculoskeletal Diseases", "Neonatal Diseases", "Nervous System Diseases", "Not Applicable", "Nutritional, Metabolic, Endocrine", "Oral Health", "Pregnancy and Childbirth", "Respiratory", "Signs and Symptoms", "Skin and Connective Tissue Diseases", "Surgery", "Urological and Genital Diseases"

#### 3.2.1.3 condition

Type: keyword (case insensitive)

The condition covered by the trial.

#### 3.2.1.4 recruitmentCountry

Type: range(string)

This matches a country in which a trial is recruiting, taken from the list below. Note that some trials may recruit in multiple countries – therefore you could match all trials which were recruiting in both France and Canada with a clause like (recruitmentCountry: "France" AND recruitmentCountry: "Canada")

"Aruba", "Afghanistan", "Angola", "Anguilla", "Aland Islands", "Albania", "Andorra", "Netherlands Antilles", "United Arab Emirates", "Argentina", "Armenia", "American Samoa", "Antarctica", "French Southern Territories", "Antigua and Barbuda", "Australia", "Austria", "Azerbaijan", "Burundi", "Belgium", "Benin", "Bonaire Saint Eustatius and Saba", "Burkina Faso", "Bangladesh", "Bulgaria", "Bahrain", "Bahamas", "Bosnia and Herzegovina", "Saint Barthelemy", "Belarus", "Belize", "Bermuda", "Bolivia", "Brazil", "Barbados", "Brunei", "Bhutan", "Bouvet Island", "Botswana", "Central African Republic", "Canada", "Cocos (Keeling) Islands", "Guernsey", "Switzerland", "Chile", "China", "Cote d'Ivoire", "Cameroon", "Congo, Democratic Republic", "Congo", "Cook Islands", "Colombia", "Comoros", "Cape Verde", "Costa Rica", "Cuba", "Curacao", "Christmas Island", "Cayman Islands", "Cyprus", "Czech Republic", "Germany", "Djibouti", "Dominica", "Denmark", "Dominican Republic", "Algeria", "Ecuador", "Egypt", "Eritrea", "Western Sahara", "Spain", "Estonia", "Ethiopia", "Finland", "Fiji", "Falkland Islands", "France", "Faroe Islands", "Micronesia, Federated States of", "Gabon", "United Kingdom", "Georgia", "Ghana", "Gibraltar", "Guinea", "Guadeloupe", "Gambia", "Guinea-Bissau", "Equatorial Guinea", "Greece", "Grenada", "Greenland", "Guatemala", "French Guiana", "Guam", "Guyana", "Hong Kong", "Heard Island and Mcdonald Islands", "Honduras",

"Croatia", "Haiti", "Hungary", "Indonesia", "Isle of Man", "India", "British Indian Ocean Territory", "Ireland", "Iran", "Iraq", "Iceland", "Israel", "Italy", "Jamaica", "Jersey", "Jordan", "Japan", "Kazakhstan", "Kyrgyzstan", "Cambodia", "Kiribati", "Saint Kitts and Nevis", "Korea, South", "Kenya", "Kuwait", "Laos", "Lebanon", "Liberia", "Libya", "Saint Lucia", "Liechtenstein", "Sri Lanka", "Lesotho", "Lithuania", "Luxembourg", "Latvia", "Macao", "Saint Martin (French part)", "Morocco", "Monaco", "Moldova", "Madagascar", "Maldives", "Mexico", "Marshall Islands", "Macedonia", "Mali", "Malta", "Myanmar", "Montenegro", "Mongolia", "Northern Mariana Islands", "Mozambique", "Mauritania", "Montserrat", "Martinique", "Mauritius", "Malawi", "Malaysia", "Mayotte", "Namibia", "New Caledonia", "Niger", "Norfolk Island", "Nigeria", "Nicaragua", "Niue", "Netherlands", "Norway", "Nepal", "Nauru", "New Zealand", "Oman", "Pakistan", "Palau", "Panama", "Pitcairn", "Peru", "Philippines", "Papua New Guinea", "Poland", "Puerto Rico", "Korea, North", "Portugal", "Paraguay", "Palestinian Territory", "French Polynesia", "Qatar", "Reunion", "Romania", "Russian Federation", "Rwanda", "Saudi Arabia", "Sudan", "Senegal", "Singapore", "South Georgia and the South Sandwich Is", "Saint Helena", "Svalbard and Jan Mayen", "Solomon Islands", "Sierra Leone", "El Salvador", "San Marino", "Somalia", "Saint Pierre and Miquelon", "Serbia", "South Sudan", "Sao Tome and Principe", "Suriname", "Slovakia", "Slovenia", "Sweden", "Swaziland", "Sint Maarten (Dutch part)", "Seychelles", "Syria", "Turks and Caicos Islands", "Chad", "Togo", "Thailand", "Tajikistan", "Tokelau", "Turkmenistan", "Timor-Leste", "Tonga", "Trinidad and Tobago", "Tunisia", "Turkey", "Tuvalu", "Taiwan", "Tanzania", "Uganda", "Ukraine", "United States Minor Outlying Islands", "Uruguay", "United States of America", "Uzbekistan", "Holy See (Vatican City State)", "Saint Vincent and the Grenadines", "Venezuela", "Virgin Islands, British", "Virgin Islands, U.S.", "Viet Nam", "Vanuatu", "Wallis and Futuna", "Samoa", "Kosovo", "Yemen", "South Africa", "Zambia", "Zimbabwe"

#### *3.2.1.5 ageRange*

Type: range(string)

Age range of participants. It will be one of the following values:

"Not Specified", "Adult", "Senior", "Neonate", "Child", "All", "Mixed", "Other"

#### *3.2.1.6 intervention*

Type: keyword (case insensitive)

The intervention being investigated in the trial.

#### *3.2.1.7 funderName*

Type: keyword (case insensitive)

Keyword search for funder name

#### *3.2.1.8 outcomeMeasures*

Type: keyword (case insensitive)

The outcome measures being investigated in the trial.

#### 3.2.1.9 inclusion

Type: keyword (case sensitive)

The inclusion criteria for trial participants.

#### 3.2.1.10 exclusion

Type: keyword (case sensitive)

The exclusion criteria for trial participants.

#### 3.2.1.11 gender

Type: range(string)

Gender of trial participants. It will be one of the following values:

"Not Specified", "Both", "Female", "Male"

#### 3.2.1.12 phase

Type: range(string)

Phase of the trial. It will be one of the following values:

"Not Specified", "Phase I", "Phase II", "Phase III", "Phase IV", "Phase I/II", "Phase II/III", "Phase III/IV", "Not Applicable"

#### 3.2.1.13 recruitmentStatus

Type: range(string)

For most trials, this is either “Not yet recruiting”, “Recruiting” or “No longer recruited”, which is computed based on the recruitment start / end dates. However, it can be overridden by an editorially set value (the same as for trialStatus 3.2.1.1 above); currently these values include “Stopped”, “Suspended” and “Enrolling by invitation”

#### 3.2.1.14 overallStartDate

Type: range(dateTime)

The start date of the trial.

#### 3.2.1.15 overallEndDate

Type: range(dateTime)

The end date of the trial, inclusive.

#### 3.2.1.16 dateApplied

Type: range(dateTime)

The date and time that the trial information was submitted to ISRCTN.

#### 3.2.1.17 lastEdited

Type: range(dateTime)

The date and time when the trial in question was last updated.

#### 3.2.1.18 sponsorOrganisation

Type: keyword (case sensitive)

The organisation sponsoring the trial; note that a trial can have multiple sponsors.

#### 3.2.1.19 fundRef

Type: keyword (case sensitive)

This allows searching for specific funders if you know their URI on the CrossRef site. For example, the Canadian Space Agency has the URI “http://dx.doi.org/10.13039/501100000016”, thus it is possible to search with a clause like (fundRef: “http://dx.doi.org/10.13039/501100000016”) to find trials with this funding source. Note that not all funders of trials in the system have an associated URI; note also that we can’t guarantee that we have the URI in the data even if it exists (for example, perhaps a funder was added to trial data before it became available in the feed from CrossRef, or maybe it has been mistyped).

#### 3.2.1.20 gridId

Type: keyword (case sensitive)

This allows searching for trial records matching a particular GRID identifier. For example, Newcastle upon Tyne Hospitals NHS Foundation Trust has GRID identifier “grid.420004.2”, so it is possible to find trials matching this foundation trust with a clause like (gridId: “grid.420004.2”). Similar caveats to fundRef apply here – it is possible that there are trials associated with an entity that should have a GRID identifier, but that GRID identifier has not been stored with the trial (for example, because of a typo in the name preventing it from being looked up in the GRID database).

#### 3.2.1.21 fundingType

Type: range(string)

Type of funding for the trial. It will be one of the following values:

"Not defined", "Charity", "Government", "Hospital/treatment centre", "Industry", "Other", "Research council", "Research organisation", "University/education"

#### 3.2.1.22 title

Type: keyword (case sensitive)

This allows searching for keywords in trial titles.

#### 3.2.1.23 acronym

Type: keyword (case sensitive)

Search the acronym field.

### 3.2.2 Querying examples

/api/query/format/default?q=covid

- ⇨ Trials containing the word “covid” somewhere in their text

/api/query/format/default?q=title:covid

- ⇨ Trials containing the word “covid” in their title

/api/query/format/default?q=title:covid AND title:pneumonia

- ⇨ Trials containing both “covid” and “pneumonia” in their title

/api/query/format/default?q=title:covid AND -title:pneumonia

- ⇨ Trials containing “covid” but not “pneumonia” in their title

/api/query/format/default?q=title:covid OR title:pneumonia

- ⇨ Trials containing “covid” or “pneumonia” (or both) in their title

/api/query/format/default?q=title:covid OR -title:pneumonia

- ⇨ Trials containing “covid” or not containing “pneumonia” in their title (be careful with queries like this; since most trials do not contain “pneumonia” this will return large number of results)

/api/query/format/default?q=title:covid AND recruitmentStatus: "Not yet recruiting"

- ⇨ Trials containing “covid” in their title, which have not yet started recruiting participants.

/api/query/format/default?q=bill and melinda and gates and foundation

- ⇨ Not very optimal way of finding trials funded by the Bill and Melinda Gates Foundation

/api/query/format/default?q="Bill and Melinda Gates Foundation"

- ⇨ Only slightly better way of finding trials funded by the Bill and Melinda Gates Foundation

/api/query/format/default?q=fundRef:"http://dx.doi.org/10.13039/100000865"

- ⇨ Best way of finding trials funded by the Bill and Melinda Gates Foundation, based on knowing the URI for this funder in CrossRef

/api/query/format/default?q=fundRef:"http://dx.doi.org/10.13039/100000865" AND ageRange:Neonate

- ⇨ Trials funded by the Bill and Melinda Gates Foundation with neonate participants

/api/query/format/default?q=fundRef:"http://dx.doi.org/10.13039/100000865" AND recruitmentCountry: "United Kingdom" AND lastEdited GE 2020-09-01T00:00:00 AND lastEdited LT 2020-11-01T00:00:00

- ⇨ Trials funded by the Bill and Melinda Gates Foundation which have the UK as one of their recruitment countries, and were last edited between 1st September 2020 and 1st November 2020

/api/query/format/default?q=recruitmentCountry GT "United States of America"

- ⇨ All trials with a recruitment country “greater than” United States of America; since this field is a range, this will match trials with at least one recruitment country that comes after that value alphabetically e.g. Yemen, Zimbabwe.

# 4 Formats

Trials data is available in 4 different formats, but we no longer recommend the use of the “internal” format. The ISRCTN system stores trial data as XML, and each output is derived from this by applying an XSLT. For the majority of people, the “default” format is likely to be the most useful; it corresponds quite closely with the actual internal format used by the system with some minor redactions (e.g. telephone numbers and email addresses where the trialists have elected to keep these private). The UKCTG and WHO formats are more specialised and are likely to be of use only to those specific entities, and there is essentially nothing in them that couldn’t be derived from the default format.

There are very minor differences depending on whether the data is the result of retrieving a single trial (3.1 above) or querying for multiple trials (3.2 above).

There are some inconsistencies with namespaces in the resulting documents which have been noted. These may be addressed in a subsequent version of the API.

## *4.1 Default*


## *4.2 UKCTG*


## *4.3 WHO*


## *4.4 Internal (deprecated)*

# 5 Parsing XML

Note that when parsing the XML downloaded from the API for further analysis, it is advised to check the output of your program, as certain characters have been known to break parsing modules.

Some example records that are known to contain these characters: 'ISRCTN16593038','ISRCTN24711056', 'ISRCTN95679074', 'ISRCTN99257265', 'ISRCTN79485429'

If using Python, the xml.etree.ElementTree module is not affected by these characters.

If the bs4.BeautifulSoup library is used in Python, then the 'html5lib' parser should be used e.g.

soup = BeautifulSoup(response.content, 'html5lib').