

News API Specification

Version:	2.3
Date Modified:	13 March 2017

1. The context.....	3
2. Function of the News API.....	3
What the module does.....	3
3. Input Parameters for API.....	4
4. Output Format for the API	5
5. Acquiring the data for API	5
6. Appendix	7
6.1 Additional Information:.....	7
6.2 Output from Data Source	9
6.3 References:.....	10

1. The context

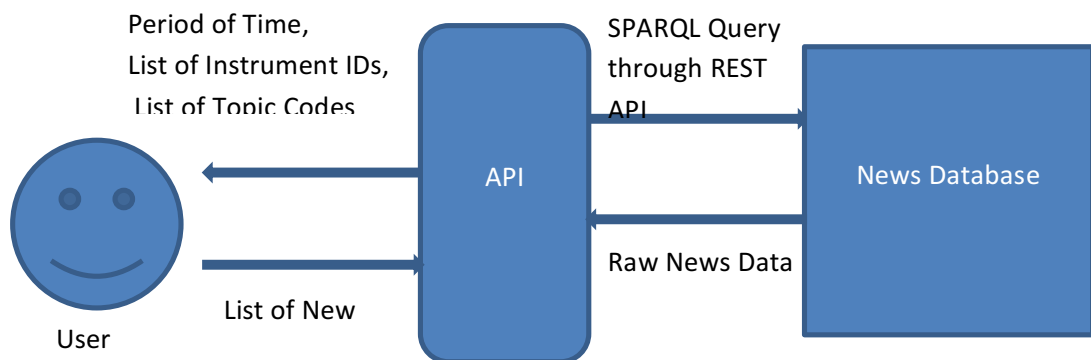
We are a company specialising in commodities trading. We are developing a complex system to provide an analytics platform for our traders. We wish to open our system to third party software houses, offering them ability to “plug-in” independent software modules that implement particular functions. We therefore request all interested companies to provide an independent software module that implements an API as specified in the rest of this document.

2. Function of the News API

This service enables the customers to receive news from all over the world. The source of these news can be Reuters or other third parties depending on your preference, if it does not affect the functionality of the API.

The objective is to build an API that would easily enable a user to find all the news related to a specific topic or for a list of specific instruments over for a given period of time.

The figure below illustrates how the API should work:



The language in which the module is to be written is not important as long as there is some way to invoke it from our system. After execution is finished, all results should be returned to users in JSON format. It is important that the module can be used without its source code being revealed.

What the module does

The module aims to isolate some specific news according to the user’s criteria.

The user has to input the 3 main information: period of interest, instrument IDs of interest , topics of interest

The module will then filter the news first according to the period (start date + end date) entered by the user and the instrument IDs and topic codes.

3. Input Parameters for API

The user has to input the 3 main information:

- period of interest
- instrument IDs of interest
- topics of interest

The Period of Time:

It's composed of two inputs:

- *A start date*
- *An end date*

Both dates have the respect the following format: **"yyyy-MM-ddTHH:mm:ss.SSSZ"**.

These inputs can NOT be empty.

For example: if I want to retrieve all the news that were issued between the first of October 2015 at 08:45:10.295 (8:45am 10 seconds and 295 milliseconds) and the first of November 2015 at 19:37:12.193 (7:37pm 12 seconds and 193 milliseconds), my inputs will be:

- `start_date = "2015-10-01T08:45:10.295Z"`
- `end_date = "2015-11-01T19:37:12.193Z"`

The list of instruments IDs:

This input contains a comma separated list of all the instruments you want to get news about. Each instrument is identified by a unique RIC (Reuters Instrument Code). *See References for more details.*

This list can be empty in the case where the user doesn't want to restrict his search.

For example: if you want to retrieve news related to ANZ bank, Woolworths Ltd, Apple Inc. and the Nasdaq Composite Index your input list should be :

- `InstrumentIDs = "ANZ.AX,WOW.AX,AAPL.O,.IXIC"`

The list of topic codes

This input contains a CSV list of all the topics you want to get news about. As for the instruments each topic has a unique identifier. *See References for the list of topics identifiers.*

This list can be empty in the case where the user doesn't want to restrict his search.

For example: if you are interested only in the news about the European Central Bank, crude oil or the aerospace sector, your input list should be:

- `TopicCodes = "ECB,CRU,AER"`

4. Output Format for the API

The module should generate and return a JSON output file to users of the following format. The “InstrumentIDs” field will return an array of instrument IDs related to each news item and “TopicCodes” field will return an array of topic codes related to each news item.

```
{
  "NewsDataSet": [
    {
      "InstrumentIDs": <>,
      "TopicCodes":<>,
      "TimeStamp": <>,
      "Headline": <>,
      "NewsText": <>
    }
    {
      "InstrumentIDs": <>,
      "TopicCodes":<>,
      "TimeStamp": <>,
      "Headline": <>,
      "NewsText": <>
    }
    .
    .
    .
  ]
}
```

Another output of the API is a log file must contain the following information:

- Developer team
- Module name and version
- Parameters passed
- An indication if execution has been successful or there is an error
- If error, indicate the nature of the error
- If successful, need to supply
 - Start date and time of execution
 - End date and time of execution
 - Elapsed time
 - Output file name

5. Acquiring the data for API

To get the news data, you can use any available data source but you are also given access to a 3Gb database that stores news obtained through NDS (Thomson Reuters News Feed Direct) and containing all the news that were published during the fourth quarter of 2015. You are provided with a REST API, through which you can query news, in SPARQL query language. Given below is a brief description on the structure of the query and how to call the database.

- Our database is hosted in a remote server and an HTTP port is open for you to run SPARQL queries on database.

URL: [http://adage.cse.unsw.edu.au:8005/v1/graphs/sparql?query="<Encoded SPARQL Query>](http://adage.cse.unsw.edu.au:8005/v1/graphs/sparql?query=)"

Observe that the HTTP service request take parameter "query". You can include any custom query as a parameter. As the query is sent via web, we need to encode the query, removing special characters.

To try this you can use an online URL encoding tool (e.g.:<http://www.url-encode-decode.com/>). Many programming languages have the facility to encode URIs.

- To test the Database API, using the *curl* command in terminal we can send an HTTP requests.

```
curl --basic --user student:studentML -i -X GET -H "Accept: application/sparql-results+json"
http://adage.cse.unsw.edu.au:8005/v1/graphs/sparql?query="<Encoded SPARQL Query>"
```

Note the parameters such as user credentials, accept type etc. we pass in the request.

```
PREFIX w3: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX fe: <http://adage.cse.unsw.edu.au/ontology/financial-events#>
PREFIX ins: <http://adage.cse.unsw.edu.au/resource/financial-events#>
PREFIX xs: <http://www.w3.org/2001/XMLSchema#>
SELECT ?s ?time ?headline ?newsBody
WHERE {
?s w3:type fe:TRTHNewsEvent.
?s fe:relatedRIC ?ric.
?s fe:timeStamp ?t.
?t fe:startTime ?time.
?s fe:newsText ?newsBody.
?s fe:headLine ?headline.
?s fe:topicCode ?topicCode.
?s fe:languageOfNews "en".
FILTER (?ric = ins:RIC_BHP.AX || ?ric = ins:RIC_BLT.L)
FILTER (?topicCode = "N2:AMERS" || ?topicCode = "N2:COM")
FILTER(xs:dateTime(?time) > "2015-10-01T00:00:00Z"^^xs:dateTime && xs:dateTime(?time) <=
"2015-10-10T00:00:00Z"^^xs:dateTime)
}
```

Given above is a SPARQL query to get news (ID, time, headline and news content) filtered by different instrument IDs, topic codes and by a period of time defined by user. Please note that each instrument code has a prefix "ins:RIC_" and each topic code as a prefix "N2" defined by the format of the database. You can have one, two or all three filters, depending on the user request.

Given below is a sample *curl* request you can run on your command line to get data from the database. The SPARQL query above is encoded as URL and appended as a parameter to the end of <http://adage.cse.unsw.edu.au:8005/v1/graphs/sparql?query=>. Copy it and paste it on your terminal to observe output. In the API you can use a library that handle HTTP requests, to access this database through REST API.

```
curl --basic --user student:studentML -i -X GET -H "Accept: application/sparql-
results+json"
http://adage.cse.unsw.edu.au:8005/v1/graphs/sparql?query="%23%23%20query%0APREFIX%20w3%3
A%20%3Chttp%3A%2F%2Fwww.w3.org%2F1999%2F02%2F22-rdf-syntax-
ns%23%3E%0APREFIX%20fe%3A%20%3Chttp%3A%2F%2Fadage.cse.unsw.edu.au%2Fontology%2Ffinancial
-
events%23%3E%0APREFIX%20ins%3A%20%3Chttp%3A%2F%2Fadage.cse.unsw.edu.au%2Fresource%2Ffina
ncial-
events%23%3E%0APREFIX%20xs%3A%20%3Chttp%3A%2F%2Fwww.w3.org%2F2001%2FXMLSchema%23%3E%0ASE
LECT%20%3Fs%20%3Ftime%20%3Fheadline%20%3FnewsBody%20%0AWHERE%20%7B%0A%3Fs%20w3%3Atype%20
fe%3ATRTHNewsEvent.%0A%3Fs%20fe%3ArelatedRIC%20%3Fric.%0A%3Fs%20fe%3AtimeStamp%20%3Ft.%0
A%3Ft%20fe%3AstartTime%20%3Ftime.%0A%3Fs%20fe%3AnewsText%20%3FnewsBody.%0A%3Fs%20fe%3Ahe
adLine%20%3Fheadline.%0A%3Fs%20fe%3AtopicCode%20%3FtopicCode.%0A%3Fs%20fe%3AlanguageOfNe
ws%20%22en%22.%0A%20%20FILTER%20(%3Fric%20%3D%20ins%3ARIC_BHP.AX%20%7C%7C%20%3Fric%20%3D
%20ins%3ARIC_BLT.L)%0A%20%20FILTER%20(%3FtopicCode%20%3D%20%22N2%3AAMERS%22%20%7C%7C%20%
3FtopicCode%20%3D%20%22N2%3ACOM%22)%0A%20%20FILTER(xs%3AdateTime(%3Ftime)%20%3E%20%22201
5-10-
01T00%3A00%3A00Z%22%5E%5Exs%3AdateTime%20%26%26%20xs%3AdateTime(%3Ftime)%20%3C%3D%20%222
015-10-10T00%3A00%3A00Z%22%5E%5Exs%3AdateTime)%0A%7D%0A"
```

The output from database will be a JSON file as shown in Appendix 6.2.

6. Appendix

6.1 Additional Information:

Teams have the choice of running their system on two different platforms:

- Standalone Program
 - ♣ PC running Windows
 - ♣ Unix/Linux platform
- Web service (accessible via a REST interface)

Throughout the workshop, each team will need to have a Web page. As a minimum, the page is showing:

- The team name and members
- Consecutive releases of their module. Each release page must include a link to download the module and information about:
 - The date and version of the release

- What has been implemented so far
- Differences with previous version
- Clear instructions on how to run the module in standalone mode
- Guidelines on how to integrate the module with other systems
- Any test software or data

6.2 Output from Data Source

```

{"head":
{"vars":["s","time","headline","newsBody"],"results":{"bindings":

{"s":{"type":"uri","value":"http://adage.cse.unsw.edu.au/resource/financial-events#20151001-183546000-nL1N1211VJ-1-2"},
"time":{"datatype":"http://www.w3.org/2001/XMLSchema#dateTime","type":"literal","value":"2015-10-01T18:35:46.961Z"},
"headline":{"type":"literal","value":"UPDATE 1-Peru copper output surges again in August on Antamina, new mines"},
"newsBody":{"type":"literal","value":"(Adds August production data for certain mines)\n LIMA, Oct 1 (Reuters) - Copper production in Peru,
the\nworld's No. 3 producer of the red metal, surged 30 percent for a\nsecond straight month in August while gold, zinc and tin output\nall
dropped, the Energy and Mines Ministry said on Thursday.\n Rising output from new mines and a 20 percent jump at\nAntamina, a top
producer, drove the copper surge in August.\nAntamina, whose 2014 output fell on weak grades, is owned by BHP\nBilliton <BHP.AX> <BLT.L>
and Glencore PLC <GLEN.L>.\n Chinalco Mining Corp's <3668.HK> Toromocho deposit, where\noutput has been ramping up, produced some
17,400 tonnes in\nAugust. Hudbay Minerals Inc's <HBM.TO> new Constancia mine\nchurched out 13,000 tonnes, official data showed.\n
Gold production was dragged down by an estimated drop in\nunregulated gold mining in Madre de Dios, the ministry said. The\ngovernment
renewed a crackdown on illegal gold mines, which sell\nore into the legal supply chain, in July. [ID:nL2NOZV11U]\n Peru was the world's third
biggest zinc, tin and silver\nproducer and seventh top gold producer last year, according to\nthe government.\n The following are official
mineral output figures for\nAugust:\n\n METAL      AUGUST      PCT CHANGE\n Copper      156,229 tonnes  +31.73\n Gold
12,298,056 ounces -3.52\n Silver      11,246,128 ounces +8.47\n Zinc      122,646 tonnes  -9.42\n Lead      25,810
tonnes  +4.64\n Iron      668,730 tonnes  +87.46\n Tin      1,693 tonnes  -5.70\n \n\n (Reporting by Mitra Taj)\n
((mitra.taj@thomsonreuters.com; +51-1-221-2130; Reuters\nMessaging: mitra.taj.thomsonreuters.com@reuters.net))\n\nKeywords: PERU
MINERALS/OUTPUT"}},

{"s":{"type":"uri","value":"http://adage.cse.unsw.edu.au/resource/financial-events#20151001-183546000-nL1N1211VJ-1-2"},
"time":{"datatype":"http://www.w3.org/2001/XMLSchema#dateTime","type":"literal","value":"2015-10-01T18:35:46.961Z"},
"headline":{"type":"literal","value":"UPDATE 1-Peru copper output surges again in August on Antamina, new mines"},
"newsBody":{"type":"literal","value":"(Adds August production data for certain mines)\n LIMA, Oct 1 (Reuters) - Copper production in Peru,
the\nworld's No. 3 producer of the red metal, surged 30 percent for a\nsecond straight month in August while gold, zinc and tin output\nall
dropped, the Energy and Mines Ministry said on Thursday.\n Rising output from new mines and a 20 percent jump at\nAntamina, a top
producer, drove the copper surge in August.\nAntamina, whose 2014 output fell on weak grades, is owned by BHP\nBilliton <BHP.AX> <BLT.L>
and Glencore PLC <GLEN.L>.\n Chinalco Mining Corp's <3668.HK> Toromocho deposit, where\noutput has been ramping up, produced some
17,400 tonnes in\nAugust. Hudbay Minerals Inc's <HBM.TO> new Constancia mine\nchurched out 13,000 tonnes, official data showed.\n
Gold production was dragged down by an estimated drop in\nunregulated gold mining in Madre de Dios, the ministry said. The\ngovernment
renewed a crackdown on illegal gold mines, which sell\nore into the legal supply chain, in July. [ID:nL2NOZV11U]\n Peru was the world's third
biggest zinc, tin and silver\nproducer and seventh top gold producer last year, according to\nthe government.\n The following are official
mineral output figures for\nAugust:\n\n METAL      AUGUST      PCT CHANGE\n Copper      156,229 tonnes  +31.73\n Gold
12,298,056 ounces -3.52\n Silver      11,246,128 ounces +8.47\n Zinc      122,646 tonnes  -9.42\n Lead      25,810
tonnes  +4.64\n Iron      668,730 tonnes  +87.46\n Tin      1,693 tonnes  -5.70\n \n\n (Reporting by Mitra Taj)\n
((mitra.taj@thomsonreuters.com; +51-1-221-2130; Reuters\nMessaging: mitra.taj.thomsonreuters.com@reuters.net))\n\nKeywords: PERU
MINERALS/OUTPUT"}},

{"s":{"type":"uri","value":"http://adage.cse.unsw.edu.au/resource/financial-events#20151001-183546000-nL1N1211VJ-1-2"},
"time":{"datatype":"http://www.w3.org/2001/XMLSchema#dateTime","type":"literal","value":"2015-10-01T18:35:46.961Z"},
"headline":{"type":"literal","value":"UPDATE 1-Peru copper output surges again in August on Antamina, new mines"},
"newsBody":{"type":"literal","value":"(Adds August production data for certain mines)\n LIMA, Oct 1 (Reuters) - Copper production in Peru,
the\nworld's No. 3 producer of the red metal, surged 30 percent for a\nsecond straight month in August while gold, zinc and tin output\nall
dropped, the Energy and Mines Ministry said on Thursday.\n Rising output from new mines and a 20 percent jump at\nAntamina, a top
producer, drove the copper surge in August.\nAntamina, whose 2014 output fell on weak grades, is owned by BHP\nBilliton <BHP.AX> <BLT.L>
and Glencore PLC <GLEN.L>.\n Chinalco Mining Corp's <3668.HK> Toromocho deposit, where\noutput has been ramping up, produced some
17,400 tonnes in\nAugust. Hudbay Minerals Inc's <HBM.TO> new Constancia mine\nchurched out 13,000 tonnes, official data showed.\n
Gold production was dragged down by an estimated drop in\nunregulated gold mining in Madre de Dios, the ministry said. The\ngovernment
renewed a crackdown on illegal gold mines, which sell\nore into the legal supply chain, in July. [ID:nL2NOZV11U]\n Peru was the world's third
biggest zinc, tin and silver\nproducer and seventh top gold producer last year, according to\nthe government.\n The following are official
mineral output figures for\nAugust:\n\n METAL      AUGUST      PCT CHANGE\n Copper      156,229 tonnes  +31.73\n Gold
12,298,056 ounces -3.52\n Silver      11,246,128 ounces +8.47\n Zinc      122,646 tonnes  -9.42\n Lead      25,810
tonnes  +4.64\n Iron      668,730 tonnes  +87.46\n Tin      1,693 tonnes  -5.70\n \n\n (Reporting by Mitra Taj)\n
((mitra.taj@thomsonreuters.com; +51-1-221-2130; Reuters\nMessaging: mitra.taj.thomsonreuters.com@reuters.net))\n\nKeywords: PERU
MINERALS/OUTPUT"}},

}}}

```

6.3 References:

What is a RIC:

<http://www.sirca.org.au/2011/01/what-is-a-ric/>

Topics Identifiers:

https://customers.reuters.com/training/trainingCRMdata/promo_content/ReutersCodes.pdf