CBK PAN 2016-10-20

ADVANCED USER GUIDE AND APPLICATION PROGRAMMING INTERFACE TO SRC SGIA$_{RV}$

AUTHORS

ŁUKASZ TOMASIK TOMASIK@CBK.WAW.PL

BEATA DZIAK-JANKOWSKA BDZIAK@CBK.WAW.PL

SUMMARY

The purpose of this document is to provide short and compact information collected from P2-SWE-I documentation and technical emails created as a support for automated data download from SRC SGIArv service.

DOCUMENT ITEMS

Summary 1

Using the ESA SSO (authorisation methods) 2

the input parameters 2

Inputs names 2

output_type parameter 3

output format of the data 3

error flag returned by the api 5

Possible latency values 6

update cadence and/or times 7

User script automatic download 7

USING THE ESA SSO (AUTHORISATION METHODS)

The SRC SGIArw is behind the ESA SSO (Open Am), so the user should first log into ESA SSO https://sso.ssa.esa.int/openam_953/UI/Login and he should have permissions to use SGIArw http://swe.cbk.waw.pl/search/select.php

After logging into SSO advanced user should keep the iPlanetDirectoryPro cookie value and use it in his requests as header input - curl 'http://swe.cbk.waw.pl/search/select.php' -H 'Cookie: iPlanetDirectoryPro=AQIC5wX2LY4SfcwGhAbCHXuKTXPPuLa0j1XtYwi7lq5dDE0.*XXXTSQAXMDE.* ' -H 'Connection: keep-alive'

Example:

curl 'http://swe.cbk.waw.pl/products/api.php?parameter=ssn&start_date=2015%2F10%2F18+00%3A00&end_date=2016%2F10%2F26+23%3A59&output_type=html' -H 'DNT: 1' -H 'Accept-Encoding: gzip, deflate, sdch' -H 'Accept-Language: pl-PL,pl;q=0.8,en-US;q=0.6,en;q=0.4' -H 'Upgrade-Insecure-Requests: 1' -H 'User-Agent: Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/54.0.2840.71 Safari/537.36' -H 'Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8' -H 'Cookie: iPlanetDirectoryPro=AQIC5wM2LY4SfcwGhAbCH0uKTXPPuLa0j1AtYwi7lq5dDE0.*AAJTSQACMDE.* ' -H 'Connection: keep-alive' --compressed

More information about using Open Am user can find at Use OpenAM RESTful Services description https://wikis.forgerock.org/confluence/display/openam/Use+OpenAM+RESTful+Services#UseOpenAMRESTfulServices-Authentication

THE INPUT PARAMETERS

Inputs names

The input parameters are send by the HTTP GET method processed by PHP script “api.php”.

The HTTP GET query body is defined by the connection of the Uniform Resource Locator (URL – commonly called site address like http://swe.cbk.waw.pl/products/api.php) and name/value pairs are joined with equal signs and different pairs are separated by the ampersand (i. e. *name1=value1&name2=value2&name3=value3),* joined by the “?” question mark;

For example: http://swe.cbk.waw.pl/products/api.php?parameter=ssn&start_date=2015/01/01 00:00&end_date=2015/12/31 23:59&output_type=html

The inputs are described below, all inputs must be present.

Input name	Input description	Possible values or format
paramerter	Index name case insensitive	ssn; f10_7; f10_7_adj; f10_7_adj_ursi; s10; m10; y10; ap; kp; dst_final; dst_provisional; ig12_final; ig12_provisional; bx_gse; by_gse; bz_gse; bmag; f30_absolute; aa_daily;
start_date	Search query start time	yyyy(/|-)mm(/|-)dd(\s|+|-)HH(:)MM i.e. 1999/01/01+00:00
end_date	Search query end time	yyyy(/|-)mm(/|-)dd(\s|+|-)HH(:)MM i.e. 1999/01/01-00:00
output_type	set output type	See table output format description

Tip 1: preferred format for date-time is yyyy-mm-dd+HH:MM i.e. 1999-01-01+00:00

Tip 2: do not mess with time separators i.e. 2015/01-01 00:00 will return 1970-01-01 00:00

Tip 3: input data are defined in http://swe.cbk.waw.pl/search/select.php

OUTPUT_TYPE PARAMETER

The output data are defined by the output_type parameter [html, CSV, JSON, PNG,PDF,EPS].

OUTPUT FORMAT OF THE DATA

value	value	output format description
html	HTML	Only for user view (2 tables, first table – query parameters+metadata+result desc, second table – results) (see fig 1).
csv	Pure CSV file (no html)	CSV file 1.Each record is located on a separate line, delimited by a line break (LF) data cell is separated by the Semicolon ';'. For example: Aaa;bbb LF Zzz;yyy LF 2.The last record in the file has an ending line break. For example: Aaa;bbb LF Zzz;yyy LF 3.The header appearing as the first line of the file starts with a hash '#'. For example: #parameter: f10_7 4. The header have information about query parameters metadata and result description (see fig 2).
json	JSON (RFC 462)	1. Data records are stored in "data" node surrounded by the "[" "]" brackets For example: "data": [["2016-10-19 17", 76.2], ["2016-10-19 20 ", 76.5]] 2. Single record is separated by the "," and date-time format is surrounded by the Quotation marks. Each record is surrounded by the "[" "]" brackets. 3. There is also extra data structure for holding query parameters metadata and columns & units description (see fig 3).

png	Portable Network Graphics PNG	1. This is a graphical representation of the result data from query The x axis is date time and y axis shows an index value 2. Plot title is on the top as description of index, the x and y labels are shown on bottom and left side off the plot 3. There is a watermark of the SRC SGIArv on the top right of the picture. 4. Data are plot as red pluses "+" marks (#ff0000) (See RFC 2083 and fig 4)
eps	PostScript graphics EPS	png converted to eps format
pdf	Portable Document Format PDF	png converted to pdf format

Tip 1: Best output for harvesting programs is json format

Tip 2: Best human readable output is csv format because of Matlab capability

Tip 3: Use http://jsonlint.com/# while developing harvester program to get human readable version

Tip 4: Use RFC 3986 for the url decoding/encoding option to send parameters with extra control characters.

Query parameters	
parameter	ssn
start_date	2015-01-01 00:00
end_date	2015-12-31 23:59
interval	default
output_type	html
Metadata	
name	Daily Total Sunspot Number
short	SSN
Results	
number of rows	365
units	,""

time	value
2015-01-01 00:00:00	104
2015-01-02 00:00:00	122
2015-01-03 00:00:00	102
2015-01-04 00:00:00	105
2015-01-05 00:00:00	88
2015-01-06 00:00:00	89

Figure 1 HTML output

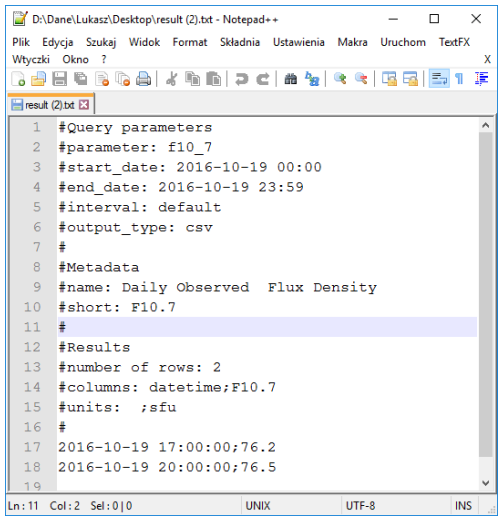

Figure 2 CSV output

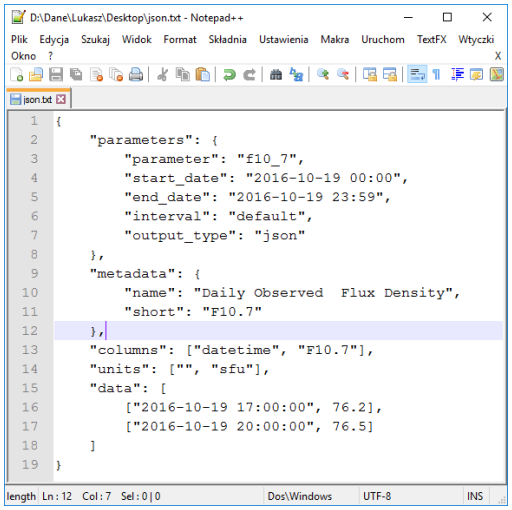

Figure 3 JSON format (human readable version)

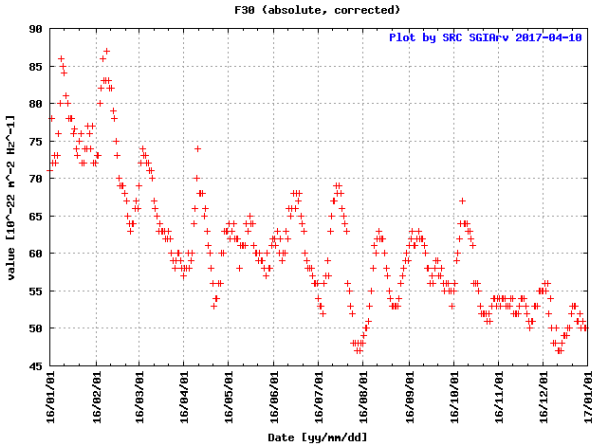

Figure 4 PNG Output

ERROR FLAG RETURNED BY THE API

The function display_html_error($error_no) in the script error_rap.php included in api.php returns the value of the error flag. Each detectable error is assigned as an alphanumeric code and symbolic name. When an error occurs, the error flag is set to the appropriate error code value. No other errors are recorded until display_html_error is called, the error code is returned, and parent script is stopped. Errors messages are show as html output.

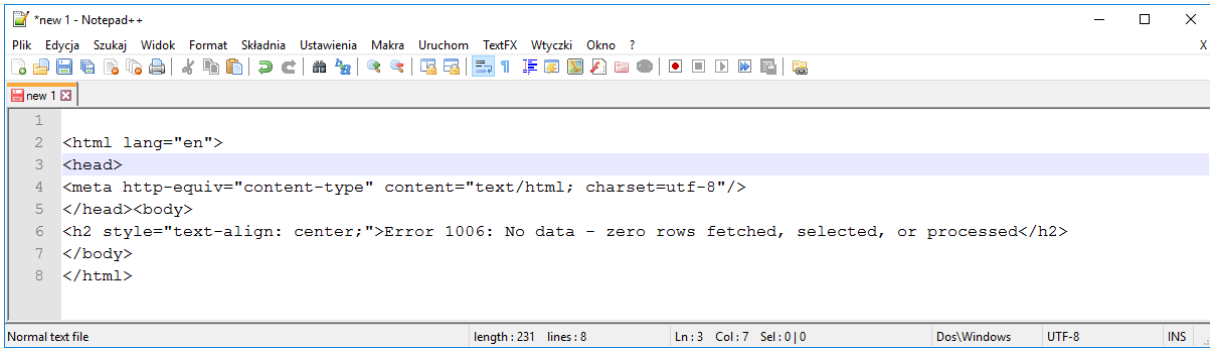

Figure 5 Sample html response in case of error detection

error #	"error message"
1001	"Error 1001: Service Temporarily Unavailable. The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later."
1002	"Error 1002: No data - parameter does not exist"
1003	"Error 1003: Maximum number of records exceeded"
1004	"Error 1004: Internal Error, Database connection failed"
1005	"Error 1005: An Unknown Error has occurred"
1006	"Error 1006: No data - zero rows fetched, selected, or processed"
1007	"Error 1007: Invalid time range"

Tip 1 The error reporting depends on output_type option it is always in HTML version except json and csv output, therefore user should seek the . \berror\b.* \b100\b[1-6].* regular expression in his output.*

Tip 2: If You use automatic harvester program based like curl and wget or any other - then for errors search ">Error 10[0-9]{2}" in given output

POSSIBLE LATENCY VALUES

Test of the beta version of PostgreSQL server has reveal that the response time is acceptable for indices with record count lower than 2Mb. Big tables (magnetic indices) have 2 days' time range limit.

Some test results for html output

Result size [number of rows]	Page load time at first run (no PostgreSQL cache) [ms]	Document Size [KB]
2	857	3,6
92	857	8,2
1190	862	64,4

7748	940	402
17591	932	897
39929	1690	2000
86393	2120	4600

Some test results for image/png output

Result size [number of rows]	Page mean load time at first run (no PostgreSQL cahe) [ms]	Document Size [KB]
2	885	5
92	902	6,5
1190	940	13,2
7748	937	13,2
17591	992	7,9
39929	1250	10,5
86393	1840	13,8

The PHP max execution time is set to 5 minutes.

TIP 1: Response latency depends also on the results size and user connection bandwidth,

UPDATE CADENCE AND/OR TIMES

All harvesting/ updating scripts run at 3:00 UTC, so the max delay form index publication and SGIArv mirroring should not exceed 24 hours.

USER SCRIPT AUTOMATIC DOWNLOAD

The automatic download can be acquired by using curl software and by implementation of the OpenAM RESTful Services API

Example:

Authentication:

curl -k 'https://sso.ssa.esa.int/openam_953/identity/authenticate?username=user&password=password'

The ESA SSO should return token:

`token.id=` AQIC5wM2LY4SfcxISzmFsu0oaa3ijnf8T0aqfzE6TapOwRs.*AAJTSQACMDE*

`This token should be passed as cookie`

iPlanetDirectoryPro=AQIC5wM2LY4SfcxISzmFsu0oaa3ijnf8T0aqfzE6TapOwRs.*AAJTSQACMDE.*

in request headers:

Curl 'http://swe.cbk.waw.pl/products/api.php?parameter=ssn&start_date=2015%2F10%2F18+00%3A00&end_date=2016%2F10%2F26+23%3A59&output_type=json' -H 'DNT: 1' -H 'Accept-Encoding: gzip, deflate, sdch' -H 'Accept-Language: pl-PL,pl;q=0.8,en-US;q=0.6,en;q=0.4' -H 'Upgrade-Insecure-Requests: 1' -H 'User-Agent: Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/54.0.2840.71 Safari/537.36' -H 'Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8' -H 'Cookie: iPlanetDirectoryPro=AQIC5wM2LY4SfcxISzmFsu0oaa3ijnf8T0aqfzE6TapOwRs.*AAJTSQACMDE.* ' -H 'Connection: keep-alive' –compressed

Reset iPlanetDirectoryPro Cookie timeout if nessery (30 minutes)

curl -k 'https://sso.ssa.esa.int/openam_953/identity/attributes?subjectid=AQIC5wM2LY4SfcxISzmFsu0oaa3ijnf8T0aqfzE6TapOwRs.*AAJTSQACMDE *&attributenames=AuthType&attributenames=Host&refresh=true' ;

Warsaw 2017-02-27