

# Introduction to REDCap & Accessing Data via API

---

OPEN STUDY GROUP

JUNE 2021

SAMMI BROWN & JESSICA LAVERY

# Outline

---



REDCap Overview



Typical process to  
export data from  
REDCap



Obtaining API access



Accessing data via API  
in R and SAS



Data manipulation

# REDCap

- Research Electronic Data Capture (REDCap) is a secure online application for creating and managing web databases
- Can collect any type of data in any environment
  - HIPAA-compliant
  - Targeted to support online and offline data capture for research studies and operations
  - Generally not used for randomized studies, but frequently used for retrospective data collection



[www.redcap.mskcc.org](http://www.redcap.mskcc.org)

# REDCap Projects

My Projects <span>Organize</span> <span>Collapse All</span> <span>Filter projects by title</span> <span>×</span> <span></span>						
Project Title		Records	Fields	Instruments	Type	Status
Unorganized Projects (1)						
Association of Exercise with COVID-19 Susceptibility and Clinical Outcomes in Patients with Cancer		970	195	15 forms	■	✓
+ Korenstein - Melanoma - IO QOL (2)						
- GENIE BPC (5)						
GENIE BPC 1.0		273	793	12 forms	■	⚙
MED19-213: Genie BPC Lung Production		1,528	853	12 forms	■	⚙
19-368: GENIE BPC - CRC Production Cohort		1,087	990	12 forms	■	⚙
19-368: GENIE BPC - BREAST Cohort		1,837	1,081	12 forms	■	⚙
19-368: GENIE BPC - Pancreas Cohort		1,367	964	12 forms	■	⚙

[Project Home](#)[Codebook](#)

The Codebook is a human-readable, read-only version of the project's Data Dictionary and serves as a quick reference for the attributes of any given field in the project without having to download and interpret the Data Dictionary. It has their coded values displayed both in the format defined by users in the Online Designer/Data Dictionary and the extended format seen in data imports and exports (i.e., field\_\_code).

[Print page](#)

#	Variable / Field Name	Field Label <i>Field Note</i>
	Instrument: <b>Curation Initiation Eligibility</b>	(curation_initiation_eligibility)
	Instrument: <b>Patient Characteristics</b>	(patient_characteristics)
	Instrument: <b>Cancer Diagnosis</b>	(cancer_diagnosis)
	Instrument: <b>Ca Directed Drugs</b>	(ca_directed_drugs)


# REDCap Instruments


- Data in REDCap are organized into instruments
- Each instrument **may** have a different data structure
  - Patient characteristics = 1 rec/patient
  - Cancer diagnosis = 1 rec/diagnosis
  - Cancer-directed drugs = 1 rec/drug regimen


# REDCap Codebook

Instrument: <b>Patient Characteristics</b> (patient_characteristics) <div>⌵ Collapse</div>																							
#	Variable / Field Name	Field Label <i>Field Note</i>	Field Attributes (Field Type, Validation, Choices, Calculations, etc.)																				
10	last_nm	Last Name PIPED IN Only <i>Piped data - do not change. Identified data that will not be shared.</i>	text, Identifier Field Annotation: PN0008																				
11	first_nm	First Name PIPED IN Only <i>Piped data - do not change. Identified data that will not be shared.</i>	text, Identifier Field Annotation: PN0009																				
12	birth_dt	Date of Birth PIPED IN Only <i>Piped data - do not change. Identified data that will not be shared. Maps to NAACCR #240.</i>	text (date_mdy, Min: 2000-01-01, Max: 2021-01-01), Required, Identifier Field Annotation: @HIDEBUTTON PN0010																				
13	birth_year	Year of Birth PIPED IN Only <i>Piped data - do not change. From GENIE data import.</i>	text (integer) Field Annotation: PN0011																				
14	adult_pedi	Patient Type: Adult or Pediatric PIPED IN Only <i>Piped data - do not change. Institutional data that will not be shared.</i>	text, Identifier Field Annotation: PN0012																				
15	naaccr_ethnicity_code	Spanish/Hispanic Origin PIPED IN Only <i>Piped data - do not change. From GENIE data import. Conforms to NAACCR #190 standards.</i>	<div>dropdown (autocomplete)</div> <table><tr><td>0</td><td>Non-Spanish; non-Hispanic</td></tr><tr><td>1</td><td>Mexican (includes Chicano)</td></tr><tr><td>2</td><td>Puerto Rican</td></tr><tr><td>3</td><td>Cuban</td></tr><tr><td>4</td><td>South or Central American (except Brazil)</td></tr><tr><td>5</td><td>Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic)</td></tr><tr><td>6</td><td>Spanish NOS or Hispanic NOS or Latino NOS</td></tr><tr><td>7</td><td>Spanish surname only</td></tr><tr><td>8</td><td>Dominican Republic</td></tr><tr><td>9</td><td>Unknown whether Spanish or not</td></tr></table>	0	Non-Spanish; non-Hispanic	1	Mexican (includes Chicano)	2	Puerto Rican	3	Cuban	4	South or Central American (except Brazil)	5	Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic)	6	Spanish NOS or Hispanic NOS or Latino NOS	7	Spanish surname only	8	Dominican Republic	9	Unknown whether Spanish or not
0	Non-Spanish; non-Hispanic																						
1	Mexican (includes Chicano)																						
2	Puerto Rican																						
3	Cuban																						
4	South or Central American (except Brazil)																						
5	Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic)																						
6	Spanish NOS or Hispanic NOS or Latino NOS																						
7	Spanish surname only																						
8	Dominican Republic																						
9	Unknown whether Spanish or not																						

# Data Entry

 **Participant Tracking**

Assign record to a Data Access Group? -- select a group -- 

 Adding new Study ID new\_subject\_test

Event Name: **Enrollment**

Study ID new\_subject\_test

**Study ID Naming Scheme**

- MSK (all sites) = S1
- LMMHC = S2
- MMC = S3
- RLCC = S4

**Study Arm**

☐ Mobile Application

☐ Informational Booklet

reset

**Did the participant receive the TEAMWork app OR informational packet?**

☐ Yes

☐ No

reset

**Intervention distribution notes**

(Please note how the randomization happened, if the patient requested another packet, etc.)

Expand

**Participant Demographics (fill out this section for ALL screened and enrolled individuals)**

**Site of Care**

**MSK Medical Record Number (MRN) or CTMS generated MRN**


# Traditional Export in REDCap

Exporting "All data (all records and fields)"


Select your export settings, which includes the export format (Excel/CSV, SAS, SPSS, R, Stata) and if you wish to perform de-identification on the data set.

Choose export format


☐

 CSV / Microsoft Excel (raw data)


☐

 CSV / Microsoft Excel (labels)


☐

 SPSS Statistical Software


☐

 SAS Statistical Software


☐

 R Statistical Software

☐

 Stata Statistical Software

☐

 CDISC ODM (XML)

De-identification options (optional)

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

Known Identifiers:

☐

Remove all tagged Identifier fields (tagged in Data Dictionary)

☐

Hash the Record ID field (converts record name to an unrecognizable value)

Free-form text:

☐

Remove unvalidated Text fields (i.e. Text fields other than dates, numbers, etc.)

☐

Remove Notes/Essay box fields

Date and datetime fields:

☐

Remove all date and datetime fields

— OR —

☐

Shift all dates by value between 0 and 364 days (shifted amount determined by algorithm for each record) [What is date shifting?](#)

[Deselect all options](#)

Advanced data formatting options

Set CSV delimiter character

Set the delimiter used to separate values in the CSV data file (only valid for CSV Raw Data and CSV Labels export formats):

, (comma) - default

Force all numbers into a specified decimal format?

You may choose to force all data values containing a decimal to have a specified decimal character (comma or period/full stop). This will be applied to all calculations and number-validated text values in the export file.

Use fields' native decimal format (default)

NOTE: Your data formatting selections above will be remembered in the future and will be pre-selected upon your next export.

Export Data

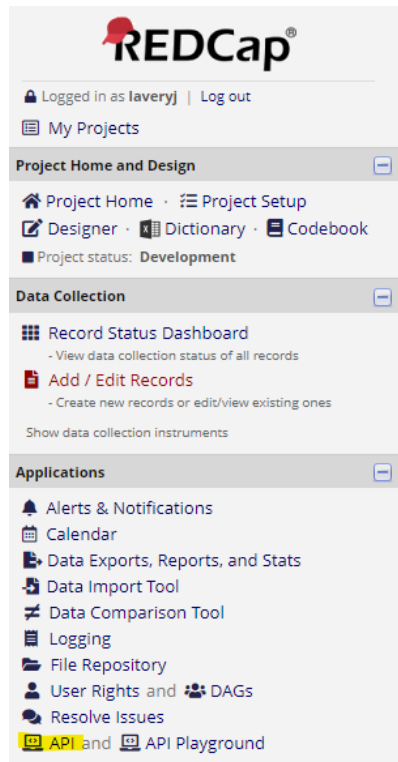
Cancel

8



# Accessing REDCap data via API

1. Email REDCap account manager for API access
2. Open project on REDCap → API → Request API token



## API

The REDCap API is an interface that allows external applications to connect to REDCap remotely, and is used for programmatically retrieving or modifying data or settings within REDCap, such as performing automated data imports/exports from a specified REDCap project. For details on the capabilities of the REDCap API and how to use it, please see the [REDCap API documentation](#).


### API Security: Best Practices

It is important to remember that when making requests to the REDCap API, you should always validate the REDCap server's SSL certificate to ensure the highest level of security during communication with the API. For details on what this means and how to do it, see the 'API Security: Best Practices' section in the [REDCap API documentation](#).

### Obtain API token for "PanCancer Prostate Training"

Use the button below to request an API token for this project from your REDCap administrator. You will need a different token for each project you would like to access. Please note that your REDCap administrator is emailed every time a token is requested.

[Request API token](#)



Logged in as laveryj | Log out

My Projects

Project Home and Design

Project Home · Codebook

Project status: Development

Data Collection

Record Status Dashboard

- View data collection status of all records

Add / Edit Records

- Create new records or edit/view existing ones

Show data collection instruments

Applications

Calendar

Data Exports, Reports, and Stats

Data Import Tool

Data Comparison Tool

Logging

File Repository

DAGs

Data Quality and Resolve Issues

API and API Playground

MED19-213: Genie BPC Lung Production

PID 1190

API

The REDCap API is an interface that allows external applications to connect to REDCap remotely, and is used for programmatically retrieving or modifying data or settings within REDCap, such as performing automated data imports/exports from a specified REDCap project. For details on the capabilities of the REDCap API and how to use it, please see the [REDCap API documentation](#).

API Security: Best Practices

It is important to remember that when making requests to the REDCap API, you should always validate the REDCap server's SSL certificate to ensure the highest level of security during communication with the API. For details on what this means and how to do it, see the 'API Security: Best Practices' section in the [REDCap API documentation](#).

Your API token for project "MED19-213: Genie BPC Lung Production"

The API token below is **ONLY for you** and will work **ONLY with this project**. This token allows special access to REDCap data and **should NOT be shared with others**. If you think your token has been compromised, then please contact your REDCap administrator immediately **AND** either delete or regenerate your token by using the buttons below.

API Token:

XXXXXXXXXX

Delete token

Finished using the API for this project? If so, please delete your token for security reasons.

Regenerate token

Think someone else knows your token? If so, please regenerate your token for security reasons.

*If someone else has your token, they can access the REDCap database. If this happens (e.g. code committed to GitHub with your token in it), you should regenerate the API token.*

10

# R Packages to Access REDCap Data via API

---

Two packages available:


1. {redcapAPI}
  - Generally works, has some trouble with special characters
2. {REDCapR}
  - Recommended R package, no issues discovered (yet!)

# {REDCapR}: redcap\_read()

```
library(REDCapR)
```

```
read_data_list= REDCapR::redcap_read(redcap_uri = "https://redcap.mskcc.org/api/",  
  token = 'YOUR_TOKEN_HERE',  
  raw_or_label = "label",  
  export_checkbox_label = TRUE)
```

```
redcap_data = as.data.frame(read_data_list$data)
```



data
success
status_codes
outcome_messages
records_collapsed
fields_collapsed
forms_collapsed
events_collapsed
filter_logic
elapsed_seconds

# {REDCapR}:

```
REDCapR::redcap_metadata_read(redcap_uri = "https://redcap.mskcc.org/api/",  
                             token = 'YOUR_TOKEN_HERE')$data,
```

	field_name	form_name	section_header	field_type	field_label
13	first_nm	patient_characteristics	NA	text	First Name<div class="red" > PIPED IN Only</div>
14	birth_dt	patient_characteristics	NA	text	Date of Birth<div class="red" > PIPED IN Only</div>
15	naaccr_sex_code	patient_characteristics	NA	dropdown	Sex<div class="red" > PIPED IN Only</div>
16	hybrid_death_dt	patient_characteristics	NA	text	Date of Death, based on NDI or other source<div class="blu...
17	ndi_cause_of_death	patient_characteristics	NA	text	Cause of Death ICD-10 Code<div class="red" > PIPED IN O...
18	last_oncvisit_dt	patient_characteristics	NA	text	Date of Last Visit to Internal Institution for Oncology Care<...
19	last_alive_dt	patient_characteristics	NA	text	Date Last Known to be Alive from Any Source<div class="bl...
20	last_anyvisit_dt	patient_characteristics	NA	text	Date of last known visit to internal Institution for any care<d...
21	enroll_hospice_dt	patient_characteristics	NA	text	Date Enrolled in Hospice Care<div class="blue" >CURATE if ...
22	primary_program_nm	patient_characteristics	NA	text	Primary Medical Oncology Program/Department<div class=...



# {REDCapR}

---

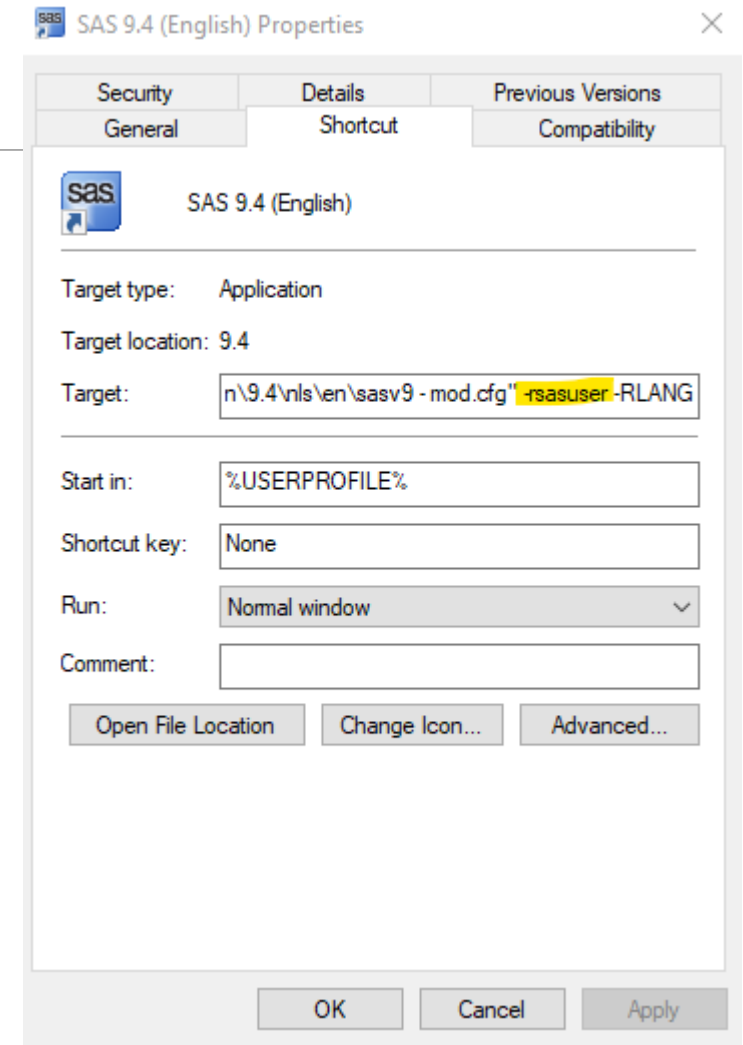
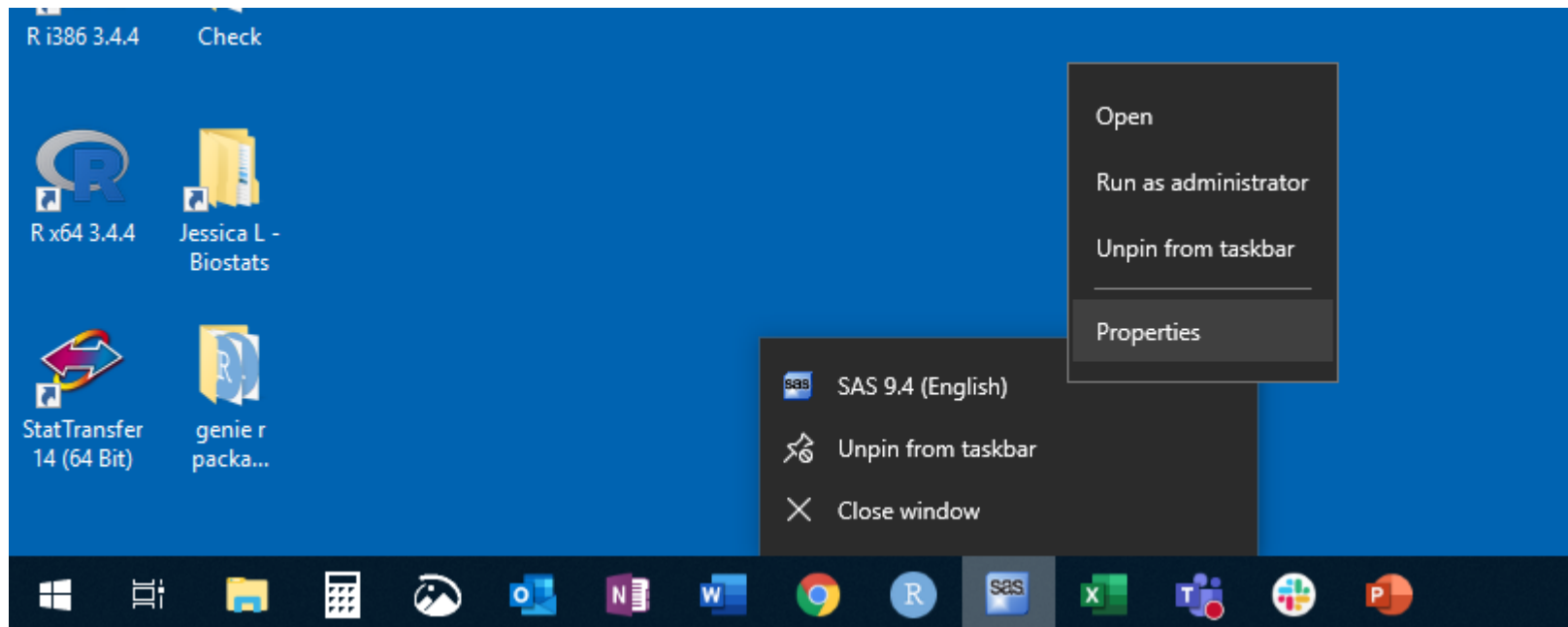
FULL DOCUMENTATION:  
[HTTPS://OUHSCBBMC.GITHUB.IO/REDCAPR/](https://ouhscbbmc.github.io/redcapR/)

# SAS Option 1: Run R Code Via SAS

---

1. Have to add “-rsasuser” system option to SAS configuration file to enable running R via SAS (Instructions on next slide)
2. Install R package (in R)
3. Run R via **PROC IML**

# Adding -rsasuser





# Run {REDCapR} Package via SAS

---

```
PROC OPTIONS OPTION=RLANG; RUN;
PROC IML;
SUBMIT/R;
  library(REDCapR)
  library(RCurl)

  read_data_list <- redcap_read(
    redcap_uri = "https://redcap.mskcc.org/api/",
    token = "XXXXXXX",
    raw_or_label = "label",
    export_checkbox_label = TRUE)

  redcap_raw_data <- as.data.frame(read_data_list$data)

ENDSUBMIT;
call ImportDataSetFromR("Work.MyData", "redcap_raw_data");
QUIT;
PROC PRINT DATA=MYDATA; RUN;
```

# SAS Option 2: PROC HTTP

---

Paper RX-02-2013

## **SAS® and REDCap API: Efficient and Reproducible Data Import and Export**

Sarah Worley, Cleveland Clinic, Cleveland, OH

Dongsheng Yang, Cleveland Clinic, Cleveland, OH

### **ABSTRACT**

REDCap (Research Electronic Data Capture), a web application for the development and use of online research databases, allows SAS® software users to download automatically-generated SAS code for importing, labeling, and formatting data. REDCap Application Programming Interface (REDCap API) provides SAS users with the option to import and export data, files, and database settings through SAS programs. The use of REDCap API with both PC and server installations of SAS 9.3 will be demonstrated.

<https://www.mwsug.org/proceedings/2013/RX/MWSUG-2013-RX02.pdf>

```

/*****
Step 1. Initial setup
*****/
* Text file for API parameters that the define the request sent to REDCap API.
Extension can be .csv, .txt, .dat;
* This file is created by this program, just specify location here;
filename my_in
"C:\Users\laveryj\Desktop\tutorial_redcap_api\sas_option2_api_parameter.txt";

* .csv output file to contain the exported data;
filename my_out "C:\Users\laveryj\Desktop\tutorial_redcap_api\redcap_raw_data.csv";

* Output file to contain PROC HTTP status information returned from REDCap API (this
is optional);
filename status
"C:\Users\laveryj\Desktop\tutorial_redcap_api\sas_option2_redcap_status.txt";

*** Project- and user-specific token obtained from REDCap ***;
%let mytoken = XXXXXX;

```

```

/*****
Step 2. Request all observations from REDCap and
write to CSV file
*****/

* Create the text file to hold the API parameters (should not need to modify);
data _null_ ;
    file my_in ;
    put "%NRStr(token=) &mytoken%NRStr(&content=record&type=flat&format=csv) &";
run;

* PROC HTTP call. Everything except HEADEROUT= is required. ***;
* this creates redcap_raw_data.csv;
proc http
    in= my_in
    out= my_out
    headerout = status
    url ="https://redcap.mskcc.org/api/"
    method="post";
run;

```

```
/******
```

```
Step 3. Read CSV file back into SAS
```

```
*****/
```

```
PROC IMPORT OUT= WORK.redcap_raw_data
```

```
    DATAFILE=
```

```
"C:\Users\laveryj\Desktop\tutorial_redcap_api\redcap_raw_data.csv"
```

```
    DBMS=CSV REPLACE;
```

```
    GETNAMES=YES;
```

```
    DATAROW=2;
```

```
RUN;
```

# Data Manipulation

- Data returned as single file (unless you pull one type of form at a time)

Record ID	Instrument	Instrument A Variable 1	Instrument A Variable 2	Instrument B Variable 1	Instrument B Variable 2
1	Patient Characteristics	Value	Value		
1	Cancer diagnosis			Value	Value
1	Cancer diagnosis			Value	Value

- Can separate the data into separate datasets (e.g. dataset for cancer diagnoses, cancer-directed drugs) by utilizing the REDCap data dictionary
- R code available on GitHub: [https://github.com/jalavery/tutorial\\_redcap\\_api](https://github.com/jalavery/tutorial_redcap_api)

# REDCap Resources

---

MSK REDCap: <https://redcap.mskcc.org/>

MSK REDCap Administrator: Andrew Zarski

REDCap training videos: <https://projectredcap.org/resources/videos/>

SAS and REDCap API: Efficient and Reproducible Data Import and Export, Worley and Yang 2013:  
<https://www.mwsug.org/proceedings/2013/RX/MWSUG-2013-RX02.pdf>

# Thank you!

---

QUESTIONS?