

# Package ‘uscoauditlog’

June 9, 2022

**Type** Package

**Title** United States Copyright Office Product Management Division SR  
Audit Data Dataset Cleaning Algorithms

**Version** 1.0.3

**Author** Frederick Liu [aut, cre]

**Maintainer** Frederick Liu <sliu85@u.rochester.edu>

## Description

Intended to be used by the United States Copyright Office Product Management Division Business Analysts. Include algorithms for the United States Copyright Office Product Management Division SR Audit Data dataset. The algorithm takes in the SR Audit Data excel file and reformat the spreadsheet such that the values and variables fit the format of the online database. Support functions in this package include `clean_str()`, which cleans instances of variable `AUDIT_LOG`; `clean_data_to_excel()`, which cleans and output the reorganized SR Audit Data dataset in excel format; `clean_data_to_dataframe()`, which cleans and stores the reorganized SR Audit Data data set to a data frame; `format_from_excel()`, which reads in the outputted excel file from the `clean_data_to_excel()` function and formats and returns the data as a dictionary that uses `FIELD` types as keys and `NON-FIELD` types as the values of those keys. `format_from_dataframe()`, which reads in the outputted data frame from the `clean_data_to_dataframe()` function and formats and returns the data as a dictionary that uses `FIELD` types as keys and `NON-FIELD` types as the values of those keys; `support_function()`, which takes in the dictionary outputted either from the `format_from_dataframe()` or `format_from_excel()` function and returns the data as a formatted data frame according to the original U.S. Copyright Office SR Audit Data online database. The main function of this package is `clean_format_all()`, which takes in an excel file and returns the formatted data into a new excel and text file according to the format from the U.S. Copyright Office SR Audit Data online database.

**License** GPL (>= 2)

**Encoding** UTF-8

**Depends** R (>= 3.4.0), stringr, openxlsx, readxl

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-06-09 08:20:02 UTC

## R topics documented:

clean_data_to_dataframe . . . . .	2
clean_data_to_excel . . . . .	3
clean_format_all . . . . .	3
clean_str . . . . .	4
format_from_dataframe . . . . .	4
format_from_excel . . . . .	5
support_function . . . . .	6

<b>Index</b>	<b>7</b>
--------------	----------

---

clean\_data\_to\_dataframe  
*Helper Function*

---

### Description

Cleans and output the reorganized SR Audit Data dataset into a data frame

### Usage

```
clean_data_to_dataframe(filename)
```

### Arguments

filename            Input name of the .xlsx file

### Value

Returns a dataframe that includes the cleaned data.

### Examples

```
## Not run:
## Read in the original excel file
filename = "data.xlsx"
clean_data_to_dataframe(filename)

## End(Not run)
```

---

clean\_data\_to\_excel     *Helper Function*

---

**Description**

Cleans and output the reorganized SR Audit Data dataset in .xlsx format

**Usage**

```
clean_data_to_excel(filename)
```

**Arguments**

filename            Input name of the .xlsx file

**Value**

Returns an excel sheet that includes the cleaned data.

**Examples**

```
## Not run:  
filename = "data.xlsx"  
clean_data_to_excel(filename)  
  
## End(Not run)
```

---

clean\_format\_all        *Main Function*

---

**Description**

Takes in a .xlsx file and returns the formatted data into a new .xlsx and .txt file according to the format of the U.S. Copyright Office SR Audit Data online database.

**Usage**

```
clean_format_all(excelfile)
```

**Arguments**

excelfile            Input the original raw SR Audit Data spreadsheet

**Value**

Returns an excel sheet and text file that includes the cleaned and formatted data that are congruent to the format of the U.S. Copyright Office SR Audit Data online database.

**Examples**

```
#This is the main function. Users should be only using this function for data cleaning.
## Not run:
filename = "data.xlsx"
clean_format_all(excelfile)

## End(Not run)
```

---

clean\_str

*Helper Function*


---

**Description**

Cleans instances of variable AUDIT\_LOG from the U.S. Copyright Office SR Audit Data spreadsheet

**Usage**

```
clean_str(str)
```

**Arguments**

str                    Input an instance value from variable AUDIT\_LOG

**Value**

Returns a cleaned string version of an instance from variable AUDIT\_LOG.

**Examples**

```
str = "2*J15*Owner2*L12*LAAS2*K10*2*C110*SR_STAT_ID2*N14*Open2*O16*Closed"
clean_str(str)
```

---

format\_from\_dataframe *Helper Function*


---

**Description**

Reads in the outputted data frame from the clean\_data\_to\_dataframe function and formats and returns the data as a dictionary that uses FIELD types as keys and NON-FIELD types as the values of those keys

**Usage**

```
format_from_dataframe(dataframedata)
```

**Arguments**

dataframedata    Input the cleaned .xlsx sheet outputted from the function clean\_data\_to\_dataframe

**Value**

Returns a vector dictionary that contains the formatted version of the cleaned data.

**Examples**

```
## Not run:
filename = "data.xlsx"
dataframedata = clean_data_to_dataframe(filename)
format_from_dataframe(dataframedata)

## End(Not run)
```

---

format\_from\_excel      *Helper Function*

---

**Description**

Reads in the outputted excel file from the clean\_data\_to\_excel function and formats and returns the data as a dictionary that uses FIELD types as keys and NON-FIELD types as the values of those keys

**Usage**

```
format_from_excel(filename)
```

**Arguments**

filename            Input the cleaned .xlsx sheet outputted from the function clean\_data\_to\_excel

**Value**

Returns a vector dictionary that contains the formatted version of the cleaned data.

**Examples**

```
## Not run:
filename = "data.xlsx"
filename = clean_data_to_excel(filename)
format_from_excel(filename)

## End(Not run)
```

---

support_function	<i>Helper Function</i>
------------------	------------------------

---

**Description**

Takes in the dictionary outputted either from the `format_from_dataframe` or `format_from_excel` function and returns the data as a formatted data frame according to the original U.S. Copyright Office SR Audit Data online database.

**Usage**

```
support_function(data)
```

**Arguments**

data	Input the dictionary variable from the <code>format_from_dataframe</code> or <code>format_from_excel</code> function
------	--

**Value**

Returns a formatted data frame according to the original U.S. Copyright Office SR Audit Data online database.

**Examples**

```
## Not run:  
filename = "data.xlsx"  
dataframedata = clean_data_to_dataframe(filename)  
data = format_from_dataframe(dataframedata)  
support_function(data)  
  
## End(Not run)
```

# Index

`clean_data_to_dataframe`, 2

`clean_data_to_excel`, 3

`clean_format_all`, 3

`clean_str`, 4

`format_from_dataframe`, 4

`format_from_excel`, 5

`support_function`, 6