# Introduction to Web Scraping with R

API Authentication


Simon Munzert | IPSDS

# API access limits

# Why API access can be restricted

- service provider wants to know who uses their API
- hosting APIs is costly—API usage limits can help control costs
- commercial interest of API hoster: you pay for access (sometimes for advanced features or massive amounts of queries only)

# Getting access to restricted APIs

## Access tokens

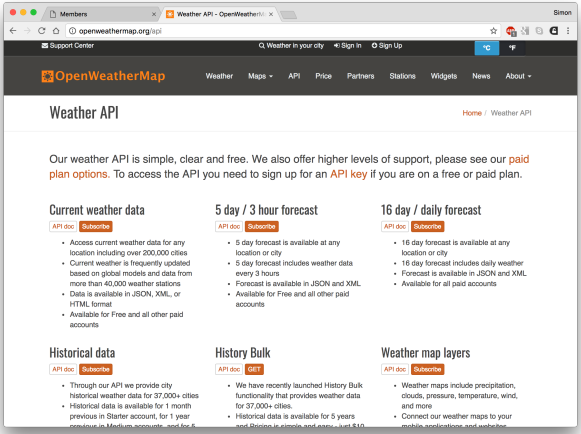
- access tokens serve as the key to the API
- they usually come in form of a randomly generated string, such as `dk5nSj485jJZP3847kjU`
- obtaining a token requires registration (often email address is sufficient)
- sometimes you have to disclose your intentions
- once you have the token, you usually pass it along with your regular API query

# Example

# Example

## The OpenWeatherMap API

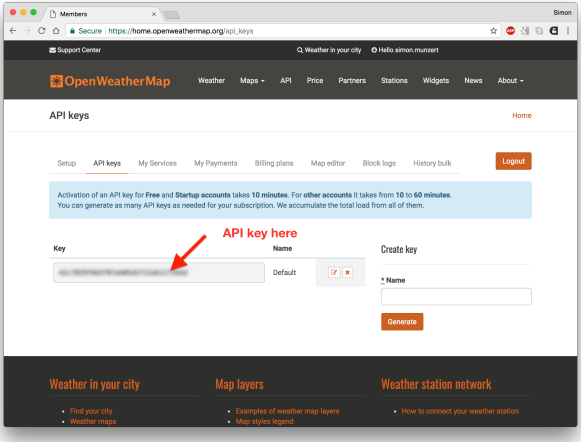
- available at http://openweathermap.org
- free access to basic weather data (paid plans for historical and more detailed data)
- sign-up necessary



# Example

## The OpenWeatherMap API

- available at http://openweathermap.org
- free access to basic weather data (paid plans for historical and more detailed data)
- sign-up necessary
- sign-up via browser—name and email address suffices



# Example

## Accessing the API from R

- copy they key to a string and store it in a local file
- import the key when you want to tap the API

R code

```
1 openweathermap <- "k4875jHkdf8n32DzZ208d7s"
2 save(openweathermap, file = "/Users/simonmunzert/rkeys.RDa")
```

end

R code

```
3 load("/Users/simonmunzert/rkeys.RDa")
4 apikey <- paste0("&appid=", openweathermap)
```

end

# Example

## Accessing the API from R

- send along the API key when you make queries to the API

R code

```
endpoint <- "http://api.openweathermap.org/data/2.5/find?"
city <- "Berlin, Germany"
metric <- "&units=metric"
url <- paste0(endpoint, "q=", city, metric, apikey)
jsonlite::fromJSON(url, flatten = TRUE)$list[1, ]
```

```
       id   name         dt rain snow                       weather
1 2950159 Berlin 1515432000   NA   NA 800, Clear, Sky is Clear, 01n
  coord.lat coord.lon main.temp main.pressure main.humidity main.temp_min
1    52.517   13.3889     -1.51          1033            54           -2
  main.temp_max wind.speed wind.deg wind.gust sys.country clouds.all
1            -1        4.1       80        NA          DE          0
```

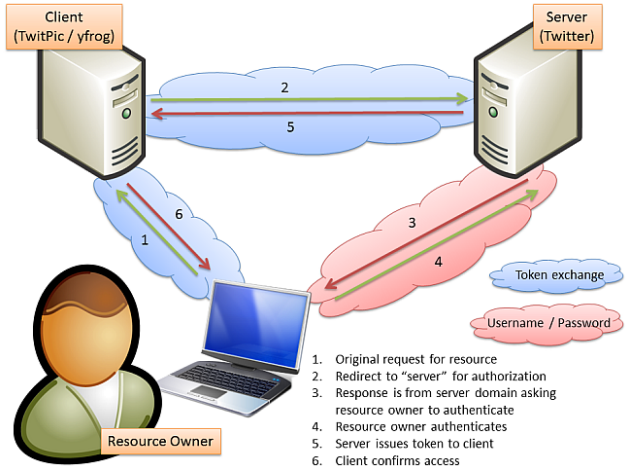
end

# OAuth authorization

# Accessing APIs with OAuth authorization

## What's OAuth?

- authorization standard
- used to provide client applications access to owner's resources—without giving them passwords
- frequently used by major companies (Twitter, Facebook Amazon, Google, ...) to allow third party applcations to interact with users accounts




Source: http://www.ubelly.com/wp-content/uploads/2010/02/OAuth1.png

# Accessing APIs with OAuth authorization

## The OAuth workflow with `httr`

- `oauth_endpoint()`: define OAuth endpoints for the request and access token
- `oauth_app()`: bundle key and secret to request access credentials
- `oauth1.0_token()` and `oauth2.0_token()`: exchange consumer key and secret for access key and secret

... but often the R API client simplifies matters (e.g., the `rtweet` package)