

# Gather Data

Date	@February 6, 2025
Assign	Terry
Status	Done
Duration(Hours)	4

informations

## informations

▼ home page: <https://www.newsinlevels.com>

▼ screen shoot

The screenshot shows the News in Levels homepage with five news items. Each item has a thumbnail, a title, a date, a brief description, and three level buttons (Level 1, Level 2, Level 3). To the right of the browser window is the Web Inspector tool, which displays the HTML and CSS code for the page. The CSS section highlights various classes and IDs used in the layout, such as 'recent-news', 'news-block', 'news-block-highlighted', and 'newsi-google-in-feed'.

data structure

```
<!-- main news block, includes 3 classes name -->
<div class="recent-news">
  <div class="news-block highlighted">
    <div class="news-block">
      <!-- ignore this class -->
      <div class="newsi-google-in-feed" id="newsi-367059809"><script async="" crossorigin="anonymous" >
        <script src="data:text/javascript;base64,IAoYWzRzYnlnb29nbGUgPSB3aW5kb3cuYWRzYnlnb29nbG
          (adsbygoogle = window.adsbygoogle || []).push({});</script>
      </div>
    </div>
  </div>
```

detail of each news

```

<!--<div class="news-block">-->
<!-- image url, includes difference size -->
<div class="img-wrap">
<a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-1/">
    
    <div class="title">
        <h3><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-1/">N
    </div>
    <!-- news date and excerpt -->
    <div class="news-excerpt">
        <p>06-02-2025 07:00</p> People in Ogoniland, Nigeria, are angry. The government wants to start oil p
    </div>
    <!-- url for 3 levels -->
    <div class="fancy-buttons">
        <ul>
            <li><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-1">L
            <li><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-2">L
            <li><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-3">L
        </ul>
    </div>
</div>

```

## pagination

```

<ul class="pagination">
    <li class="active">
        <span class="current">1</span></li>
    <li>
        <a href="https://www.newslevels.com/page/2/" class="inactive">2</a></li>
    <li>
        <a href="https://www.newslevels.com/page/3/" class="inactive">3</a></li><li><a href="https://www.ne
    </ul>

```

▼ sub page(sample): <https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-1/>

▼ screen shoot

The screenshot shows a news article titled "Nigeria wants to produce oil again – level 1". The article is dated 06-02-2025 07:00. It features a small image of oil platforms, some text about oil production in Ogoniland, and a video player for a cleanup project. The Web Inspector on the right displays the DOM tree and various CSS rules applied to the elements.

## data structure

```
<div class="article">
<!--this is the main content-->
<div class="article-upper">
<!--this is the audio block-->
<div class="article-lower">
<!-- skip elements-->
<iframe src='video link'>
</div>
```

### ▼ content detail

#### 1. get title

```
$$('.article-title')[0].innerText
```

'Nigeria wants to produce oil again – level 1'

#### 2. image

```
$$('.img-wrap')[0].children[0].href
```

<https://www.newslevels.com/wp-content/uploads/2025/02/Novy-projekt-4.jpg>

#### 3. body

```
let strings = $$('#nContent')[0].children;
let strings_len = strings.length;
let strings_body = '';
let strings_date = '';
let strings_words = '';

// first element is date
// last two elements are 'difficult words' and 'level tip'
for(var i = 1;i<strings_len-2;i++){
    strings_body+=strings[i].innerText+'\n';
```

```

}
console.log(strings_body+'\n');

```

```

> let strings = $$('#nContent')[0].children;
let strings_len= strings.length;
let strings_body = '';
let strings_date = '';
let strings_words='';

// first element is date
// last two elements are 'difficult words' and 'level tip'
for(var i = 1;i<strings_len-2;i++){
    strings_body+=strings[i].innerText+'\n';
}
console.log(strings_body, '\n');

Protesters in Nigeria oppose the government's plan to restart oil production in Ogoniland, one of the VM2579:12
most polluted regions in the world. For decades, oil spills have poisoned Ogoniland's water and land, making
life difficult for its people.
A 1 billion dollar cleanup project launched in 2018 has failed, leaving residents frustrated and demanding more
effort and compensation. Despite the massive budget, cleanup efforts remain ineffective. Many had hoped oil
production would bring prosperity, but instead, they lost their clean environment. The cleanup agency HYPREP
faces serious allegations of mismanagement. Its leader denies corruption, but Nigerian authorities have warned
he could face arrest.
As the government moves to restart oil extraction, many fear a repeat of past environmental destruction.
Families of activists who lost their lives opposing oil exploitation in 1995 believe their struggle was in vain.
For many in Ogoniland, suffering has replaced the prosperity they once expected.

```

#### 4. learning words

```

strings_words = strings[strings_len-2];
let difficult_words = new Map();
let strings_words_nodes = strings_words.childNodes;

for(var i=0; i<strings_words_nodes.length;i++){
    if (strings_words_nodes[i].childElementCount>0){
        difficult_words.set(strings_words_nodes[i].text.trim(),strings_words_nodes[i+1].textContent.trim().replace(/\n/g));
    }
}
difficult_words.forEach((val,key)=>{
    console.log(`Word: ${key}, Desc.: ${val}`);
})

```

```

> strings_words = strings[strings_len-2];
let difficult_words = new Map();
let strings_words_nodes = strings_words.childNodes;

for(var i=0; i<strings_words_nodes.length;i++){
    if (strings_words_nodes[i].childElementCount>0){
        strings_words_nodes[i].text

        } difficult_words.set(strings_words_nodes[i].text.trim(),strings_words_nodes[i+1].textContent.trim());
    }

}
difficult_words.forEach((val,key)=>{
    console.log(`Word: ${key}, Desc.: ${val}`);
})
Word: compensation, Desc.: (money paid to someone in exchange for something that has been lost or for VM5379:18
some problem),
Word: mismanagement, Desc.: (managing something badly or wrongly), VM5379:18
Word: in vain, Desc.: (without success or a result), VM5379:18
Word: exploitation, Desc.: (making use of and benefiting from resources). VM5379:18

```

raw data:

```

<!--  <div class="article-upper"> -->

<div class="article-title">
<h2>Nigeria wants to produce oil again – level 1</h2>

```

```

</div>
<div class="upper-content">
<!-- we can ignore this-->
<div class="img-wrap">
<a href="https://www.newslevels.com/wp-content/uploads/2025/02/Novy-projekt-4.jpg" class="fancybox i

<!-- we can ignore this-->
<div class="level fancy-buttons">
<ul>
<li class="active"><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-l
<li><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-2">Level :
<li><a href="https://www.newslevels.com/products/nigeria-wants-to-produce-oil-again-level-3">Level :
</ul>
</div>
</div>
<div id="nContent">
<p>06-02-2025 07:00</p>
<p>People in Ogoniland, Nigeria, are angry. The government wants to start oil <strong>production</strong> a
<p>Oil production makes water in Ogoniland dirty. The land is not safe for farmers. A <strong>cleanup</strong>
<p>Now, people want real help and money for their <strong>losses</strong>. Long ago, people think that oil w
<p>Difficult words: <a href="https://www.newslevels.com/?s=production" target="_blank" rel="noopener">
    <a href="https://www.newslevels.com/?s=production" target="_blank" rel="noopener">
        <strong>production</strong></a> (making something), <a href="https://www.newslevels.com/?s=clea
<p>You can watch the original video in the Level 3 section.</p>
</div>

<!-- Go to www.addthis.com/dashboard to customize your tools --> <div class="addthis_native_toolbox"></div>
</div>

```

## ▼ video detail

video link

1. `video url` get the element from sub page

```

$$(‘iframe’)[0]
$$(‘iframe’)[0].src
//https://w.soundcloud.com/player/?url=https%3A//api.soundcloud.com/tracks/2024831296&color=%23ff

```

2. http get `resolve`

url: <https://api-widget.soundcloud.com/resolve?>

```

fetch("https://api-widget.soundcloud.com/resolve?url=https%3A//api.soundcloud.com/tracks/2024831296&color=%23ff
  "headers": {
    "accept": "application/json, text/javascript, */*; q=0.01",
    "accept-language": "en-US,en;q=0.9,zh-CN;q=0.8,zh-TW;q=0.7,zh;q=0.6,ja;q=0.5",
    "sec-ch-ua": "\"Not A(Brand)\";v=\"8\", \"Chromium\";v=\"132\", \"Google Chrome\";v=\"132\"\"",
    "sec-ch-ua-mobile": "?0",
    "sec-ch-ua-platform": "\"macOS\"",
    "sec-fetch-dest": "empty",
    "sec-fetch-mode": "cors",
  }

```

```

    "sec-fetch-site": "same-site"
},
"referrer": "https://w.soundcloud.com/",
"referrerPolicy": "strict-origin-when-cross-origin",
"body": null,
"method": "GET",
"mode": "cors",
"credentials": "omit"
});

```

1. parameters

- a. **url:** <https://api.soundcloud.com/tracks/2024831296>
- b. **format:** json
- c. **client\_id:** gqKBMSuBw5rbN9rDRYPqKNvF17ovlObu
- d. **app\_version:** 1738665620

2. parameters get from:

- a. url → from page element ifram
- b. client\_id → from widgetxxx.js, just provide two options, can http get the result and search the **keyword with underline**  
<https://widget.sndcdn.com/widget-9-fc25f73c498c.js>

```
client_id: u ? "nljtjiYnjkOhMyh5xrbqEW12DxeJVnic" : "gqKBMSuBw5rbN9rDRYPqKNvF17ovlObu",
```

- c. app\_version → also from the same widgetxxx.js

```

e((function() {
    t.start("1738665620")
})

```

<https://api-widget.soundcloud.com/resolve?url=https%3A//api.soundcloud.com/tracks/2024831296&format=json>

2. http get **tracks**

url: <https://api-widget.soundcloud.com/users/1060338772/tracks?>

```

fetch("https://api-widget.soundcloud.com/users/1060338772/tracks?limit=20&offset=0&linked_partitioning=1",
"headers": {
    "accept": "application/json, text/javascript, */*; q=0.01",
    "accept-language": "en-US,en;q=0.9,zh-CN;q=0.8,zh-TW;q=0.7,zh;q=0.6,ja;q=0.5",
    "sec-ch-ua": "\"Not A(Brand\";v=\"8\", \"Chromium\";v=\"132\", \"Google Chrome\";v=\"132\")",
    "sec-ch-ua-mobile": "?0",
    "sec-ch-ua-platform": "\"macOS\"",
    "sec-fetch-dest": "empty",
    "sec-fetch-mode": "cors",
    "sec-fetch-site": "same-site"
},
"referrer": "https://w.soundcloud.com/",
"referrerPolicy": "strict-origin-when-cross-origin",
"body": null,
"method": "GET",
"mode": "cors",

```

```
    "credentials": "omit"
});
```

1. parameters
  - a. **limit:** 20
  - b. **offset:** 0
  - c. **linked\_partitioning:** 1
  - d. **format:** json
  - e. **client\_id:** gqKBMSuBw5rbN9rDRYPqKNvF17ovlObu
  - f. **app\_version:** 1738665620

find out the `hls` url , there are two possible string. get the first element

```
//regex:  
https://api-widget.soundcloud.com/media/soundcloud:tracks:2024831296/.*?/stream/hls
```

2. http get `hls`

url: `https://api-widget.soundcloud.com/media/soundcloud:tracks:2024831296/792cd406-05a3-44a7-b237-fd926d88b04a/stream/hls?`

```
fetch("https://api-widget.soundcloud.com/media/soundcloud:tracks:2024831296/792cd406-05a3-44a7-b237-fd926d88b04a/stream/hls?  
    "headers": {  
        "accept": "*/*",  
        "accept-language": "en-US,en;q=0.9,zh-CN;q=0.8,zh-TW;q=0.7,zh;q=0.6,ja;q=0.5",  
        "sec-ch-ua": "\"Not A(Brand\";v=\"8\", \"Chromium\";v=\"132\", \"Google Chrome\";v=\"132\")",  
        "sec-ch-ua-mobile": "?0",  
        "sec-ch-ua-platform": "\"macOS\"",  
        "sec-fetch-dest": "empty",  
        "sec-fetch-mode": "cors",  
        "sec-fetch-site": "same-site"  
    },  
    "referrer": "https://w.soundcloud.com/",  
    "referrerPolicy": "strict-origin-when-cross-origin",  
    "body": null,  
    "method": "GET",  
    "mode": "cors",  
    "credentials": "omit"  
};
```

1. parameters

- a. **client\_id:** gqKBMSuBw5rbN9rDRYPqKNvF17ovlObu

respond, this link can download for m3u8 file

```
{  
    "url": "https://cf-hls-media.sndcdn.com/playlist/i94SpoPIOruk.128.mp3/playlist.m3u8?Policy=eyJ...  
}
```

3. http get `playlist`

url: `https://cf-hls-media.sndcdn.com/playlist/i94SpoPIOruk.128.mp3/playlist.m3u8?`

```
fetch("https://cf-hls-media.sndcdn.com/playlist/i94SpoPIOruk.128.mp3/playlist.m3u8?Policy=eyJ...  
    "headers": {  
        "accept": "*/*",  
        "accept-language": "en-US,en;q=0.9,zh-CN;q=0.8,zh-TW;q=0.7,zh;q=0.6,ja;q=0.5",  
    }
```

```

    "priority": "u=1, i",
    "sec-ch-ua": "\"Not A(Brand\";v=\"8\", \"Chromium\";v=\"132\", \"Google Chrome\";v=\"132\")",
    "sec-ch-ua-mobile": "?0",
    "sec-ch-ua-platform": "\"macOS\"",
    "sec-fetch-dest": "empty",
    "sec-fetch-mode": "cors",
    "sec-fetch-site": "cross-site"
},
"referrer": "https://w.soundcloud.com/",
"referrerPolicy": "strict-origin-when-cross-origin",
"body": null,
"method": "GET",
"mode": "cors",
"credentials": "omit"
});

```

1. parameters

a. **Policy:**

eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I

b. **Signature:** ANy3ILSU3HDqXzm5hT9mJlp-EiPG0H1R6yB2IKJntM1HPWdeiFZkh7MT4DxVmC-

1EkP6yzw2a6-p3kjKZNMMi9JDQ7qxHS-ov-  
0nQ6AXUbPYNAEDrXtm43XvKEdIE4p6ROx~EnGKuOkEzMgnWcfZoaJHKCY-  
OgKPXpTnW2VJ6p5izbCKL3Fa4-ILgL7EkPP4O6IQVJis3xSSCanLKpZ1TN6q~q0onOMCm-  
tEK9c5KxpGUee~cexiHZSBj9n3QznDbv~RmUmbnjd1E24rm2~KwT17dqho4sIDLbXGOfdqGuppZpcL\

c. **Key-Pair-Id:** APKAI6TU7MMXM5DG6EPQ

2. parameters get from

all parameters → from `hls` respond

respond, URL for **each segmented** audio /color

```

#EXTM3U
#EXT-X-VERSION:6
#EXT-X-PLAYLIST-TYPE:VOD
#EXT-X-TARGETDURATION:10
#EXT-X-MEDIA-SEQUENCE:0
#EXTINF:1.985272,
https://cf-hls-media.sndcdn.com/media/159660/0/31762/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:2.977908,
https://cf-hls-media.sndcdn.com/media/159660/31763/79410/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:4.989302,
https://cf-hls-media.sndcdn.com/media/159660/79411/159240/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/159241/318900/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/318901/478561/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/478562/638221/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/638222/797882/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/797883/957542/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:9.978604,
https://cf-hls-media.sndcdn.com/media/159660/957543/1117202/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXTINF:5.198278,
https://cf-hls-media.sndcdn.com/media/159660/1117203/1200376/i94SpoPIOruk.128.mp3?Policy=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJkZWZlNjM1Y29tL3BsYXIsaXN0I
#EXT-X-ENDLIST

```

## conclusion

I can now get the parsed data from the web, the next step is to use Swift code to implement it and save to local storage. the UI will start after that.