

# More RSS feeds

November 20, 2020

## Quick links

- <https://seanhly.github.io/rss/gov.ie.xml>
- [https://europa.eu/newsroom/highlights.xml\\_en](https://europa.eu/newsroom/highlights.xml_en)
- [https://europa.eu/newsroom/highlights.xml\\_fr](https://europa.eu/newsroom/highlights.xml_fr)
- [https://europa.eu/newsroom/highlights.xml\\_de](https://europa.eu/newsroom/highlights.xml_de)

## What am I talking about?

The world is a frustrating place at the moment, and reading the news too much doesn't help. I've curated a small collection of RSS feeds that are informative to Irish people and that avoid information overload.

The first feed relates to local government news. We are all citizens, and the most important news to a citizen is what their government is doing. 'Government' involves many sub-departments and offices, and the news feed I chose reflects this. This is the source of my Irish government news feed. This feed pulls from many government department press releases and publications, ordered by date. I then converted this webpage to an RSS feed using the below scripts:

## The BASH script

```
#!/usr/bin/bash
# URL linked earlier
url='https://www.gov.ie/en/...'
itemSelector='body>div.container>div.row>div>div.row>div>ul>li'
start() {
```

```

        dst="$(find -L -maxdepth 5 -name gov.ie.xml\
| awk '{print length($0)"\t"$0}'\
| sort -n\
| head -1\
| cut -d$'\t' -f2)"
        if [ ! "$dst" ]; then
dst="$HOME/Documents/Homepage/remote/rss/gov.ie.xml"
mkdir -p "$(dirname "$dst")"
        fi
        echo "$dst"
        curl "$url"\
| hxnormalize -l 240 -x\
| hxselect -s '\n' -c "$itemSelector"\
| awk -f "$script" > "$dst"
}
script="$(dirname $0)/gov.ie.awk"
(cd "$HOME" && start)

```

## The AWK script

```

BEGIN {
    title = "gov.ie - Publications"
    description = "Find publications from government\
departments and agencies, including news, press releases,\
reports, and Freedom of Information responses. This page is\
still in development and is not a complete list of all publications.\
Further information on publications can be found on the individual\
department websites."
    language = "en"
    print "<?xml version='1.0' encoding='utf-8'?>"
    print "<channel>"
    print "\t<title>"title"</title>"
    print "\t<description>"description"</description>"
    print "\t<language>"language"</language>"
}
/^\s*<a / {
    match($0, /^\s*<a[>]+href="([~"]*)" [^>]*>([~<]*)/, array)
    print "\t<item>"
    print "\t\t<title>"array[2]"</title>"
}

```

```

    print "\t\t<link>https://www.gov.ie"array[1]"</link>"
}
/^\s*<p>/ {
    match($0, /\s*<p>\s*([\^;]*)\s*([\^;]*)\s*([\^;]*)/; array)
    print "\t\t<description>"array[3]"</description>"
    print "\t\t<pubDate>"array[1]"</pubDate>"
    print "\t\t<author>"array[2]"</author>"
    print "\t</item>"
}
END {
    print "</channel>"
}

```

The above scripts run automatically from my various machines each day, using `cron`, and the generated news feed is located here (<https://seanhly.github.io/rss/gov.ie.xml>).

Now, the second source of news is from the broader EU newsroom. I really love their RSS feeds, because of two things:

1. They only have about one news item per day
2. They publish their news in many European languages

This addresses the information overload problem, while also giving readers the opportunity to brush up their foreign language skills.

Here are some of their feeds:

- English
- French
- German