



How (and why) to build your own Pricewatch(ish) (web)application

Import, Slice & Dice, Present!



About me

- Donovan van der Roest
- Matando @ tweakers.net
- 17 years + experience in web (app) development
- Programmer, geek, wine lover



Overview

- Introduction
- Why create your own importer
- Use cases
- How to create your own
- Pricewatch overview



Introduction

**Building my own Pricewatch?
Feeling all Jedi now?**



Import driven affiliate application





import driven affiliate application?

An application built to promote products or services from partner shops for which they will reward you after any sale or referral.

The pricewatch is an import driven web application.



Affiliate?



Do I need to use an affiliate network?

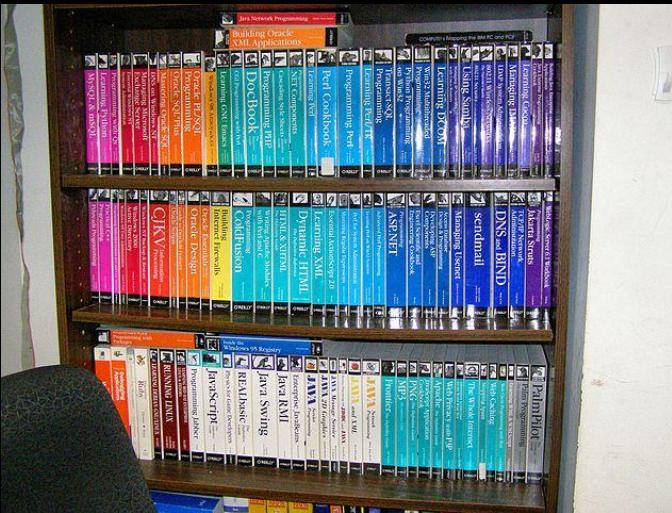


Why would you create your own importer?



Why? It's a good exercise

- You learn a lot by developing your own importer
- You drive the project, choose to do the parts you like





Why? It's fun

- Get involved in a business area you care about
- Learn more about the product / service yourself
- Help others to find this cool product or service





Why?

A nice alternative for advertisement driven sites

- Banner revenue is low (when not blocked)
 - Product / service click-through is more effective





Use cases



Use cases Tweakers Pricewatch

The screenshot shows the Tweakers Pricewatch homepage. At the top, there's a navigation bar with links for Nieuws, Reviews, Pricewatch, Vraag & Aanbod, Forum, and Meer. Below the navigation is a search bar with the placeholder "Zoek een product" and a "MATANDO" logo. The main content area features a sidebar with "Populaire categorieën" and a grid of product categories with icons and names: Smartphones, Laptops, Televisies, Videokaarten, Monitoren; Kabels, Processors, Moederborden, Koptelefoons, Games; Tablets, Solid state drives, Geheugen intern, Koelkasten, Behuizingen. At the bottom, there are buttons for "Help andere Tweakers, geef je mening!", "Product beoordelen", and "Winkel beoordelen". A section titled "Producten met de grootste prijsdaling" shows thumbnail images of various products.

Added value

- Single place to look for technical products & more
- Easy to find your desired product
- Product & store reviews help shoppers



Use cases ViniFinder.nl

Gemakkelijk wijn vinden
Vul een zoekterm in

Landen Soorten Druiven Aanbiedingen

wijn uit Spanje

Op deze pagina vind je 673 flessen wijn uit Spanje die direct online te koop zijn in Nederland

Verfijn selectie op: Type Regio Wijnhuis Jaar Druif

Sorteer: Meest bekeken

Acantus Golden Selection Bo... 2015 De 2013 Rioja Reserva is bekroond met 2 zilveren medailles en is koningin van Bodegas Olarra: 2 jaar achter elkaar uitgeroepen tot beste producent van Spanje en zelfs één van de 20 beste wijnhuizen ter wereld. Dat prof je ... Meer over deze wijn	Olarra Clasico Reserva a Rioja 2016 Een loepzuivere Sauvignon Blanc die al met 2 keer zilver en 1 keer goud is bekroond! Bodegas Olarra is één van de 20 beste wijnhuizen ter wereld en de nummer 1 van Spanje! De witte wijn bevat de frisheid van Santerre en het ... Meer over deze wijn	Acantus Sauvignon Blanc Bod... 2016 Sappig, fluweelzacht en ronig van smaak met ladingen zoete kersen, elkenhout, cigarbox en vanille. Na 24 maanden houtrijping en ruim 6 jaar in fles is deze fenomenale Gran Reserva in topvorm! De Spaanse rode wijn komt van ... Meer over deze wijn	Dominios del Marquesado Gra... 2009 Vina Agustina Catalunya DO ... 2011 Fluwelzachte Gran Reserva uit CatalunyaAls er een wijn is uit Catalunya die de prachtige Rioja is, dan is het de fluweelzachte Gran Reserva van Vina Agustina. Heerlijk voor de verfijnde wijnliefhebber! Groots, krachtig, ... Meer over deze wijn	Verdejo 2016 EL PARADOR De Parador wijnen worden gemaakt van door de Chítito familie die maken wijnen in Rioja, Navarra en Rueda. Hun wijnen zijn kenmerkt zich door veel fruit in de smaak. Zo ook in deze Verdejo, de inheemse druif uit Rueda. Prof ... Meer over deze wijn	Don Mendo Edición Limitada ... 2011 Zeldzame super cuvée van drie kroonjaren van Bodegas Don Mendo. De Spaanse rode kroonjatzer is gemaakt van de moeilijke Cabernet uit 2009, de molligste Merlot uit 2010 en de kruidige Syrah uit 2011. Het resultaat is een ... Meer over deze wijn	Casa Safra Gran Reserva Sel... 2011 Zeldzame 6 jaar oude Gran ReservaDe Casa Safra Limited Edition Ovaal Gran Reserva werd oorspronkelijk alleen voor de privé collectie gemaakt en ook uitsluitend geschonken aan vrienden en familie van de ... Meer over deze wijn
--	--	--	--	---	--	--

Added value

- A single place to find almost all wines sold in the Netherlands and Belgium
- Easy for visitors to filter down on to the desired product



Use cases Trivago.nl

The screenshot shows the Trivago.nl search interface. The search bar at the top contains the query "Toscane, Pension, Vakantiehuis, Casa Rural, 4 sterren". Below the search bar are filters for arrival date (zaterdag, 01-09-18) and departure date (donderdag, 06-09-18). The main results section displays three vacation rental options:

- Montebelli Agriturismo & Co...** (4 stars)
Gavorrano, Toscane, Italië
Priced from €116 to €426.
Details: Casa Rural, Als beste gewaardeerd, 252 reviews, Gratis wifi.
- Fattoria Pratale** (4 stars)
Tavarnelle Val di Pesa, Toscane, Italië
Priced from €153 to €128.
Details: Casa Rural, Goed (78 reviews), Gratis wifi.
- Streda Belvedere** (4 stars)
Vinci, Toscane, Italië
Priced from €112 to €102.
Details: Casa Rural, Uitstekend (137 reviews), Gratis wifi.

On the right side of the results, there is a sidebar with a "Sorteer op" dropdown set to "Alleen populariteit" and a "Zie deal" button. The sidebar also lists prices from various travel websites like Hotels.com, roomXXL.nl, Expedia, and Booking.com, along with their respective deal prices.

Added value

- A single place to look for hotels
- Compares the comparison sites

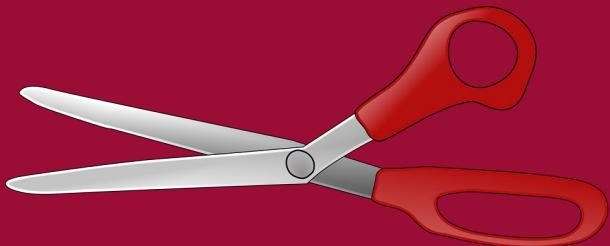


Create your own



Preparation

Getting set-up

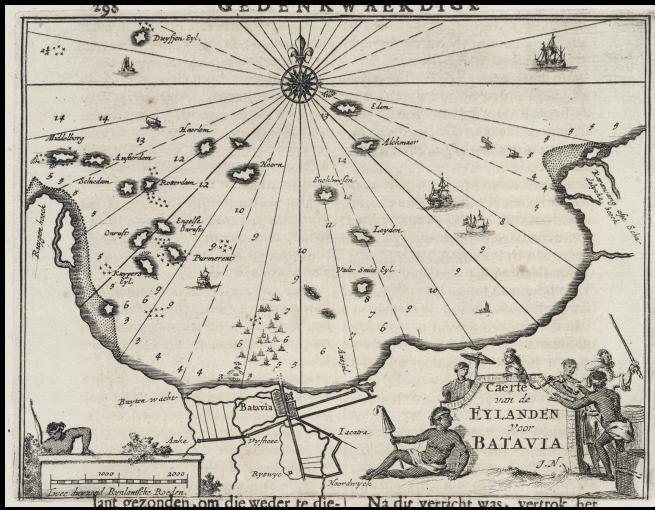




Preparation

Find a market

- Find a market you are passionate about
- Research that market for competitors
- Check if you can easily add merchants





Preparation

Consider carefully

- Consider ease of purchase online
- It's easier to sell posters online than monster trucks





Preparation

Find networks / partners to work with

Affiliate networks



FAMILY
BLEND

Partner programs





Create your own

Tips and tricks for writing your Importer



Tips and tricks for writing your Importer

Create a blueprint

- Start with a generic abstract class
- Think about which methods to override in advance



```
abstract class AbstractProductImporter implements ItemImporter{

    private $itemRepository;

    public function __construct(ItemRepository $itemRepository){
        $this->itemRepository = $itemRepository;
    }

    abstract protected function isItemInStock($feedNode): bool;

    private function fetchItemByUrl(string $url): Item {
        $item = $this->itemRepository->findOneBy($url);
        if($item !== null) {
            return $item;
        }

        $item = new Item();
        $item->setUrl($url);
        $item->setAvailable(true);

        return $item;
    }
}
```



Tips and tricks for writing your Importer

Only extend when absolutely needed



```
/**
 * Returns if an item is in stock
 *
 * @param Object $feedNode
 *
 * @return bool
 */
protected function isItemInStock($feedNode): bool
{
    $stock = $this->readFeedPropertyValue($feedNode, 'stock');
    if( empty($stock) ){
        return false;
    } elseif ( $stock[0] > 0 ) {
        return true;
    }
    return false;
}
```



Tips and tricks for writing your Importer

Only extend when absolutely needed



```
/**  
 * AbstractProductImporter.php  
 */  
protected function getStockValue($feedNode) {  
    $stock = $this->readFeedPropertyValue($feedNode, 'stock');  
    if( empty($stock) ){  
        return false;  
    }  
    return $stock;  
}  
  
/**  
 * DaisyconImporter.php  
 */  
protected function isItemInStock($feedNode): bool  
{  
    $stock = $this->getStockValue($feedNode);  
    return isset($stock[0]) && $stock[0] > 0;  
}
```



Tips and tricks for writing your Importer

Think abstract

- Think abstract
- Use configuration data
- Create a data mapper



Tips and tricks for writing your Importer

Think abstract

```
mapping:  
    descriptionLong: description  
    descriptionShort: description  
    country: subsubcategories  
    deliveryTime: deliveryTime  
    winetype: categoryPath  
    url: URL  
    name: name  
basics:  
    id: ID  
configuration:  
    removeWhiteSpaceFromImage: false  
    alwaysInStock: false  
    minDescriptionLength: 70
```



Tips and tricks for writing your Importer

Think abstract



```
protected function mapProperties(){
    $affiliateMappingData = $this->getAffiliateMappingData();
    foreach($affiliateMappingData as $key => $value){
        $string = $this->getPropertyValue($this->feedWineNode,$value);
        if($string !== null && $string !== ''){
            $this->setWineProperty($key,$string,true);
        }
    }
}
```



Tips and tricks for writing your Importer

No stopping here

- The show must go on!



Tips and tricks for writing your Importer

No stopping here



```
if (empty($wine->getName())) {  
    throw new Exception('Skipping wine: no name');  
}  
if (strlen($wine->getLongDescription()) < $this->minDescriptionLength) {  
    throw new Exception('Skipping wine: description too short');  
}  
if ($wine->getCountry() === null || $wine->getWineType() === null) {  
    throw new Exception('Skipping wine: no mapping possible');  
}
```



Tips and tricks for writing your Importer

Spread the effort

- Offload time consuming tasks to separate task with multiple threads



Tips and tricks for writing your Importer

Spread the effort



```
// Used to pre-warm cache before running the actual import
$skipCacheWarmup = Validator::validateBoolean($input->getOption(self::OPTION_SKIP_CACHE_WARMUP));
if ( ! $skipCacheWarmup) {
    $removeXmlCache = $input->getOption(self::OPTION_CLEAR_CACHE);
    $icecatImporter->warmupDownloadCache($groupedIcecatIndexProducts, $removeXmlCache);
}
```



Tips and tricks for writing your Importer

About threads

- Threads allow you to run the same task multiple times in parallel.
- Useful especially for simple time consuming tasks such as downloading files



```
public function executeWithMultipleThreads(array $productsList, int $threads = 4): void {
    $processes = [];

    for ($i = 0; $i < $this->threads; $i++) {
        $pid = pcntl_fork();

        if ($pid === -1) {
            throw new RuntimeException('Failed to start child process');
        }

        if ($pid > 0) {
            $processes[] = $pid;
        } else {
            foreach ($productsList[$i] as $product) {
                $this->productService->addProductImages($product);
            }
            exit;
        }
    }

    foreach ($processes as $pid) {
        pcntl_waitpid($pid, $status);
    }
}
```



Tips and tricks for writing your importer

Don't trust anyone

- Never trust external data
- Sanitise everything
- Verify files to be what they claim



Tips and tricks for writing your Importer

Sanitise data

```
public function sanitise($value) {
    if(is_int($value)){
        return $this->sanitiseInt($value);
    }
    return $this->sanitiseString($value);
}

private function sanitiseString(string $value) {
    $sanitisedValue = filter_var($value, FILTER_SANITIZE_STRING);
    return $sanitisedValue ?? '';
}

private function sanitiseInt(int $value) {
    return (int) $value;
}
```



Tips and tricks for writing your Importer

Sanitise Images



```
public function isSupportedImage($imagePath){  
    $supportedImageTypes = [  
        IMAGETYPE_GIF,  
        IMAGETYPE_JPEG,  
        IMAGETYPE_PNG  
    ];  
  
    $imgType = exif_imagetype($imagePath);  
  
    return in_array($imgType, $supportedImageTypes, true);  
}
```



Tips and tricks for writing your Importer

Enhance

- Missing data? Ask, the merchant might add it for you
- Use external services / APIs to enhance supplied data
- Do not mix API calls within your import scripts / run afterwards

Any  API

 ebay developers program

 facebook for developers

 Developer

 ProgrammableWeb

 LinkedIn Developers



Tips and tricks for writing your Importer

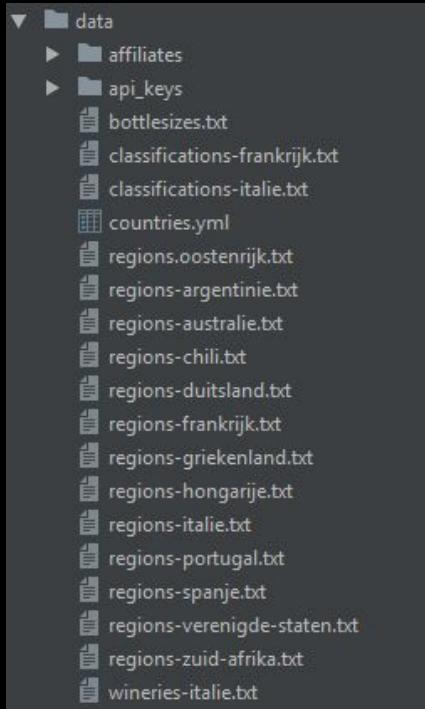
Enhance

- Still missing? Add your own



Tips and tricks for writing your Importer

Enhance



countries:

-
name: Argentinië

alternatives:

- Argentinie
- Argentina

-
name: Australië

alternatives:

- Australie
- Australia

-
name: Chili

-
name: Duitsland

alternatives:

- Germany

```
public function getMatchingCountryFromString(string $countryName): ?string {
    $countries = $this->getYamlFromFile('countries.yml');

    if(empty($countries)) {
        return null;
    }

    foreach($countries as $country){

        if(strpos($country['name'], $countryName) !== false){
            return $country['name'];
        }
        if(isset($country['alternatives'])){

            foreach($country['alternatives'] as $alternative){

                if(strpos($alternative, $countryName) !== false){
                    return $country['name'];
                }
            }
        }
    }
    return null;
}
```



Create your own Present



Present

Show off your end-product

- Build for growth
- Use caching!
- But don't forget to flush ;-)





This is how we roll

A pricewatch overview



Tweakers Pricewatch

Technical setup

- PHP 7.2
- Symfony 3 (soon 4)
- Mysql 5.6
- Other than Symfony complete custom built
- Jenkins to schedule



Tweakers Pricewatch

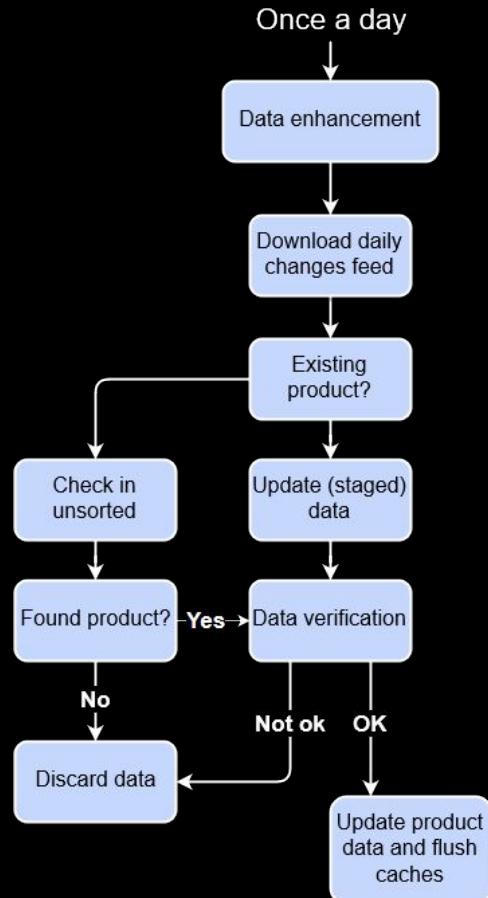
Metrics

- 425.000+ products
- 985.000+ prices
- 32.000+ reviews
- +- 320 shops
- 18 import moments p/d
- 100GB+ data import p/d



Tweakers Pricewatch

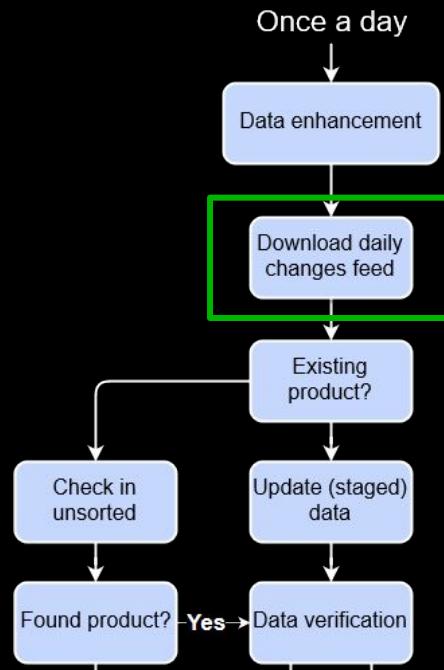
Import flow





Tweakers Pricewatch

Specific features



- Download Daily update XML
- Identify which products need to be updated or added
- Download product XML files using 10 threads before the main import



Tweakers Pricewatch

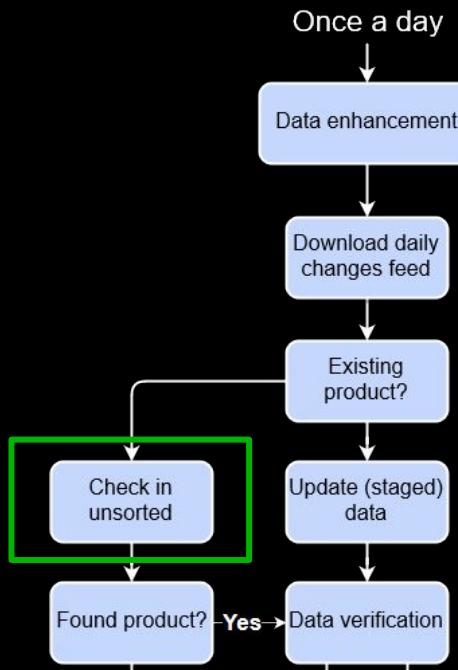
Specific features





Tweakers Pricewatch

Specific features

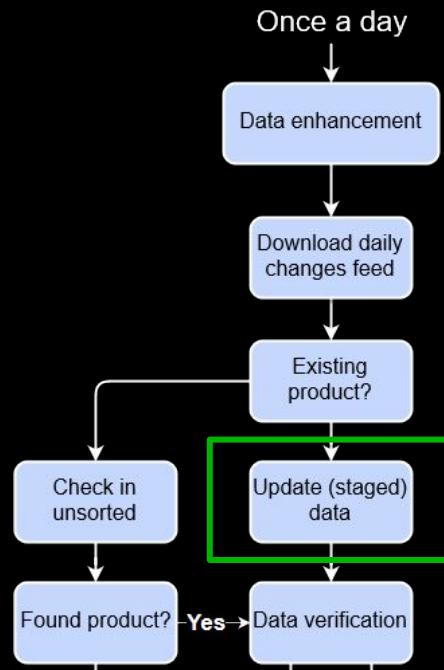


- Check unsorted products to find matching products (on ean, sku)
- Add found products, discard others



Tweakers Pricewatch

Specific features

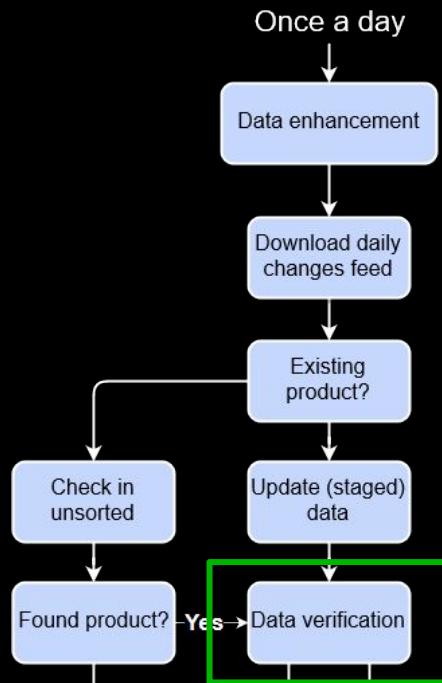


- Product details are added / updated:
 - Specifications
 - Images
 - Product codes (ean, sku)



Tweakers Pricewatch

Specific features



- Some categories are verified automatically (e.g. cables)
- Others are verified manually
- Others are ignored



Tweakers Pricewatch

Data verification

Afbeeldingen

Afbeeldingen No files selected.

	2001335947.jpeg	<input checked="" type="radio"/> Primair	Verwijder
	2001335951.jpeg	<input type="radio"/>	Verwijder

Specificaties

Dit is de primaire uitvoering, alle andere uitvoeringen nemen standaard de specificaties met de waarden van deze uitvoering over.
label (x) = er zijn x secundaire uitvoeringen met een andere waarde

Schermdiagonaal <input checked="" type="checkbox"/>	<input type="text" value="24"/> "
Gebogen scherm <input checked="" type="checkbox"/>	<input checked="" type="radio"/> Leeg <input type="radio"/> Ja <input type="radio"/> Nee
Beeldschermcoating <input checked="" type="checkbox"/>	<input type="radio"/> Leeg <input type="radio"/> Glanzend <input checked="" type="radio"/> Mat <input type="radio"/> Onbekend
Resolutie <input checked="" type="checkbox"/>	was: 1920x1080 (Full HD) <input type="text" value="1920x1200 (WUXGA)"/> <input type="button" value="▼"/>
Dot-pitch <input checked="" type="checkbox"/>	<input type="text" value="0,27"/> mm
Soort paneel <input checked="" type="checkbox"/>	<input type="radio"/> Leeg <input type="radio"/> E-ink <input checked="" type="radio"/> LCD <input type="radio"/> OLED
LCD-paneel <input checked="" type="checkbox"/>	<input type="text" value="IPS"/> <input type="button" value="▼"/>



Tweakers Pricewatch

Future challenges

- Further automation of data verification
- Continue to optimize import efficiency



Recap



**Go out there & build something you
would love!**

Questions?





Contact me

Donovan van der Roest



<https://tweakers.net/gallery/18603/>



<https://nl.linkedin.com/in/donovan-van-der-roest>



<https://github.com/dvdroest/import-driven-affiliate-application-presentation>