Implementing a FOSS4G Welsh language base map in the Azure Cloud for the Welsh Government

Sam Franklin   FOSS4G UK   2022-11-17

CGI

Overview

- Continuation from Mapio Cymru project and requirements
- Proof of Concept and challenges
- Deploy to production
- Tuning
- Go Live

The Goal

Add the Mapio Cymru base map as an option to the Data Map Wales Map Viewer.
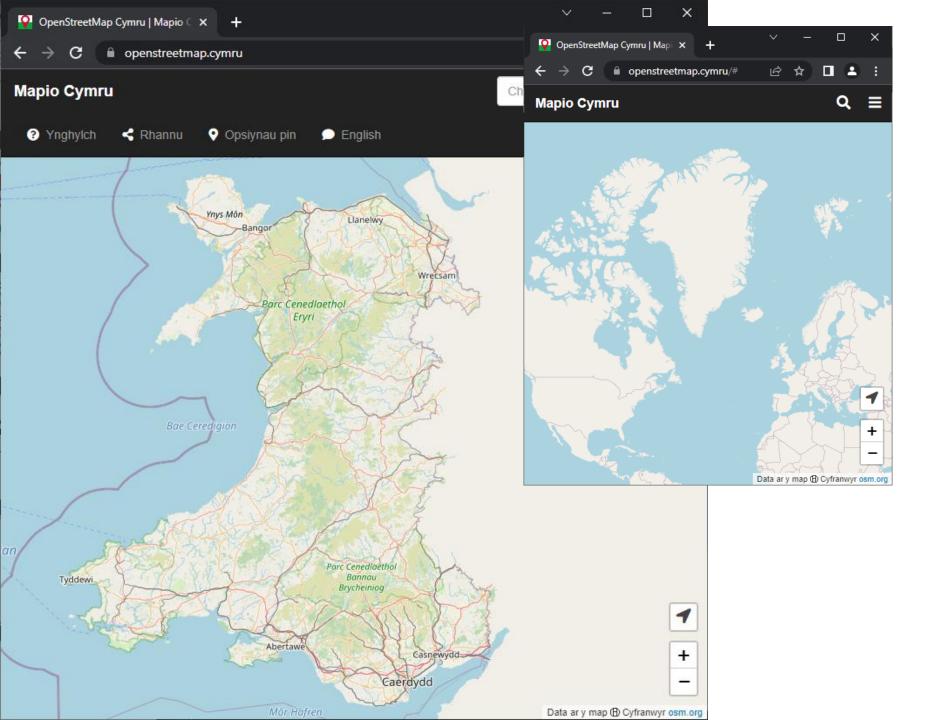

Project Stakeholders

Welsh Government
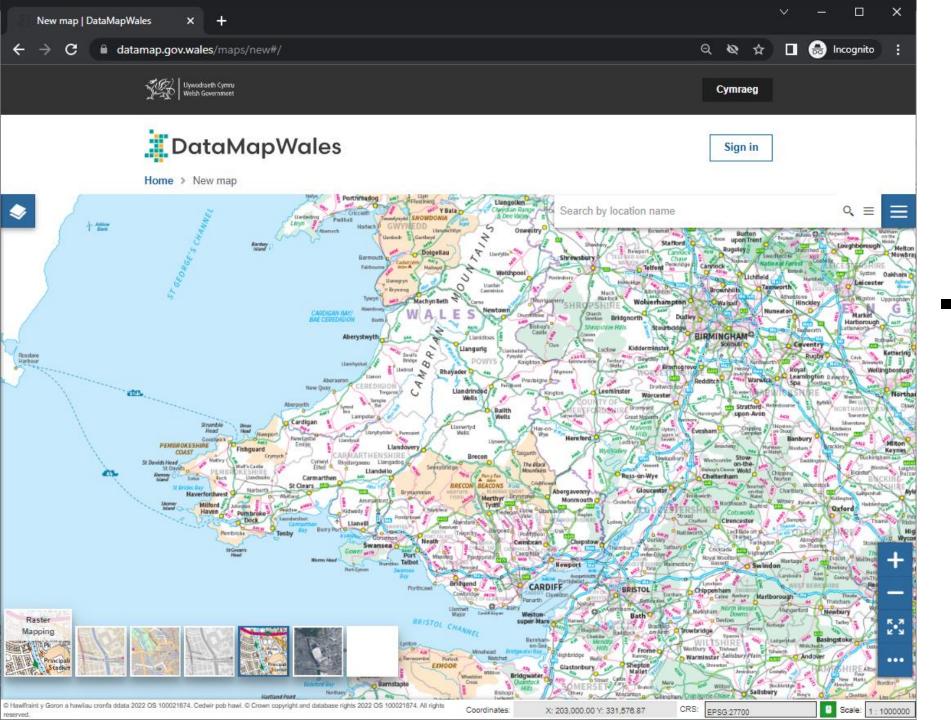

Delivery partner for Data Map Wales

Mapio Cymru

Just add the Mapio Cymru map tiles into that Map Viewer?

= Web Mercator

= Web Mercator

.. we want British National Grid

Check the requirements!

Requirements
Source Welsh Language place names from Mapio Cymru
Supports British National Grid projection
Supports OGC Web Map Tiled Service (WMTS)
Self-hosted
Maintain familiar Openstreetmap styling

Solution

Requirements	
Source Welsh Language place names from Mapio Cymru	✓
Supports British National Grid projection	✓
Supports OGC Web Map Tiled Service (WMTS)	✓
Self-hosted	✓
Maintain familiar Openstreetmap styling	?

PostgreSQL

GEOSERVER   GeoWebCache

MapStore 2

DataMapWales

Styling Geoserver Layers like OpenStreetMap.org

Mapnik !=

mapnik combines pixel-perfect image output with lightning-fast cartographic algorithms, and exposes interfaces in C++, Python, and Node.

CartoCSS

CartoCSS (short: Carto) is a language for map design. It is similar in syntax to CSS, but builds upon it with specific abilities to filter map data and by providing things like variables. It targets the Mapnik renderer and is able to generate Mapnik XML or a JSON variant of Mapnik XML. It can run from the command line or in the browser by using a bundler like Browserify.

Mapnik !=

OSM Styles

Mapnik !=

OSM Styles

Dockerise + vanilla data

Mapnik !=

OSM Styles

Dockerise + vanilla data

English = na!

Requirements	
Source Welsh Language place names from Mapio Cymru	❌
Supports British National Grid projection	✅
Supports OGC Web Map Tiled Service (WMTS)	✅
Self-hosted	✅
Maintain familiar Openstreetmap styling	✅
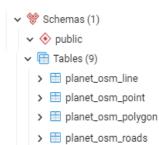
Should be an easy swap of the database aaaaand …

osm2pgsql

imposm

Schemas (1)
- public
  - Tables (9)
    - planet_osm_line
    - planet_osm_point
    - planet_osm_polygon
    - planet_osm_roads

Database

OpenStreetMap Carto uses as its standard database PostgreSQL with PostGIS geospatial extension and is named "gis". It is derived from the OSM database, but does not store the full dataset – only the rendering-related elements are available. Since v4.0.0, OpenStreetMap Carto uses the PostgreSQL hstore (extension and data type) and is pre-processed with a lua script by the osm2pgsql tool.

Switch2OSM
Take back control of your maps

Tables (27)
- osm_admin
- osm_amenities
- osm_barrierpoints
- osm_barrierways
- osm_boundary
- osm_buildings
- osm_housenumbers
- osm_housenumbers_interpolated
- osm_landusages
- osm_landusages_gen0
- osm_landusages_gen1
- osm_landusages_gen2
- osm_landusages_gen3
- osm_places
- osm_roads
- osm_roads_gen0
- osm_roads_gen1
- osm_roads_gen2
- osm_roads_gen3
- osm_transport_areas
- osm_transport_points
- osm_waterareas
- osm_waterareas_gen0
- osm_waterareas_gen1
- osm_waterways
- osm_waterways_gen0
- osm_waterways_gen1

README.md

OSM Bright styles

Introduction

This is a GeoServer data directory with OSM layer and styles mimicking a OSM-bright style.

README.md

Also, the tool used to parse, reclassify and import the data into PostgreSQL, called imposm, works only on Linux. We have prepared a command line tool that can be run on Linux to prepare a PostgreSQL database dump, which can then be moved to a Windows machine and restored.

With this in mind:

1. Download a Imposm binary as linked from the documentation site

2. Download the preferred OSM extract from the Geofabrik download server, in pbf format.

Public

Is there a workaround?

Data Flow / Components

- [✓] Uses Mapio Cymru data
- [✓] BNG projection
- [✓] WMTS
- [✓] OSM Styling

Imposm importer loads data to postgres in imposm schema format

GeoFabrik PBF file → ETL A → Postgres db (imposm schema) → Updated PG database with Mapio Cymru data (imposm schema) → App Service: Geoserver + Geowebcache OSM Style SLDs → DMW Map Viewer

Mapio Cymru PG dump file (osm2pgsql schema) → ETL B → Postgres db (osm2pgSQL schema) → ETL C → Updated PG database with Mapio Cymru data (imposm schema)

pg restore loads Mapio Cymru data to Postgres in osm2pgsql schema format

Custom SQL to generate a lookup of "osm_id" and "name" from Mapio Cymru. Lookup table then JOIN + UPDATE to the IMPOSM schema tables which is compatible with the Geoserver stack.

Public

and to production

Azure
app service

DataMapWales

Geoserver +
GWC
deployed as docker
image in an Azure
App Service

**Tiles as blobs
In Azure Storage
Account**

**OSM vector data in
dbs in PG Host**

**Container images in
private registry**

docker

Azure
app service

Parameterised Config

App Service

Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Microsoft Defender for Cloud
- Events (preview)

Deployment

- Quickstart
- Deployment credentials
- Deployment slots
- Deployment Center

Settings

- Configuration
- Authentication
- Application Insights
- Identity
- Backups
- Custom domains
- Custom domains (preview)
- TLS/SSL settings
- Certificates (preview)
- Networking

Configuration ☆ ⋯

↻ Refresh 🖫 Save ✕ Discard ♡ Leave Feedback

Application settings   General settings   Path mappings

Application settings

Application settings are encrypted at rest and transmitted over an encrypted channel. You can choo as environment variables for access by your application at runtime. Learn more

➕ New application setting   👁 Show values   ✎ Advanced edit

🔽 Filter application settings

Name	Value	S
COMMUNITY_EXTENSIONS	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
	👁 Hidden value. Click to show value	A
JAVA_OPTS	👁 Hidden value. Click to show value	A

Environment Variable Configuration

Azure app service

Parameterised Config

Scale out

DataMapWales

Invisible
Load
Balancer

Scale out
instances

**Tiles as blobs
In Azure Storage
Account**

**OSM vector data in
dbs in PG Host**

**Container images in
private registry**

Azure
app service

Parameterised
Config

Scale
out

Azure
DevOps

Azure DevOps

Pipelines

Recent   All   Runs

Recently run pipelines

Pipeline	Last run
✅ data Load	#20220804.1 • added test suit Manually triggered for T
✅ Update data artifacts (geofabrik and Mapi...	#20220804.1 • added test suit Manually triggered for T
✅ Update webapp	#20220620.2 • add seeding sc Manually triggered for SF
✅ Build custom geoserver image	#20220620.2 • add seeding sc Manually triggered for SF
✅ Delete tile cache only	#20220520.4 • removed redur Manually triggered for SF
✅ Create webapp	#20220519.4 • fix for paramet Manually triggered for SF
✅ Create postgres environment	#20220519.1 • updated docke Manually triggered for SF
✅ Delete webapp	#20220518.3 • fix failing pipeli Manually triggered for SF
✅ Delete postgres environment	#20220517.1 • add data looku Manually triggered for SF

Example Data Update Workflow

Update artifacts

#1 Initiate Pipeline

Run pipeline
Select parameters below and manually run the pipeline ✕

Branch/tag
🔀 develop ▾
Select the branch, commit, or tag

Environment
◉ dev
○ prd ──────────────── Select dev or prd

☐ Update geofabrik package ──────────────── Select Geofabrik (this is a PBF format file)

☐ Update mapiocymru package ──────────────── Select MapioCymru (this is a postgres dump file)

Enter MapioCymru blob name (assumes in the root of the container)
mapiocymru.dump ──────────────── Enter the filename of the MapioCymru blob in the storage account

#2 Pipeline jobs run tasks

Update Geofabrik Azure Artifact

∨ ✅ Download and publish geofabrik | 42s
  ⊘ Initialize job | 2s
  ✅ Checkout MapioCymruOGCSta... | 1s
  ✅ Download geofabrik dataset w... | 18s
  ✅ Publish geofabrik artifact to az... | 19s
  ✅ Post-job: Checkout MapioCy... | <1s
  ⊘ Finalize Job | <1s

Update MapioCymru Azure Artifact

∨ ✅ Download and publish MapioCy... | 50s
  ⊘ Initialize job | 3s
  ⊙ Pre-job: Get Key Vault Secrets | <1s
  ✅ Checkout MapioCymruOGCSta... | 1s
  ✅ Get Key Vault Secrets | <1s
  ✅ Download MapioCyrmu datase... | 9s
  ✅ Publish MapioCymru data arti... | 34s
  ✅ Post-job: Checkout MapioCy... | <1s
  ⊘ Finalize Job | <1s

#3 Versioned data artifacts stored in Azure DevOps

/ Artifacts / Feeds

osm-build-artifacts ▾

▽ Filter by keywords

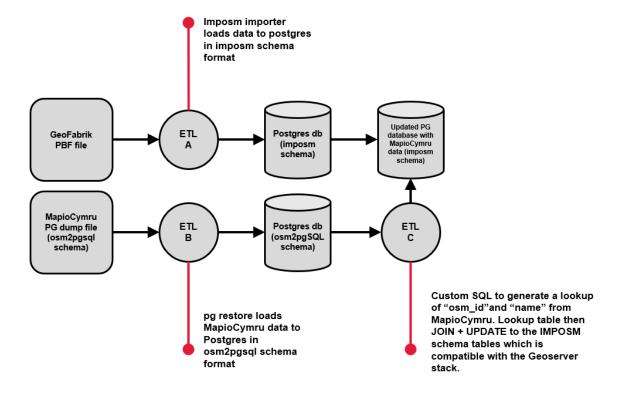|	Type	Package	Views	Source	Last Pushed
☐	🗄	dmw-osm-low-res-gpkg Version 0.0.1		This feed	3 May
☐	🗄	geofabrik-wales-latest-osm-pbf Version 0.0.7		This feed	18 May
☐	🗄	imposm Version 0.11.1		This feed	11 May
☐	🗄	mapiocymru-latest Version 0.0.6		This feed	18 May
☐	🗄	test-package Version 0.0.1		This feed	3 May

Update artifacts

Load artifacts to databases

Load data to db

∨ ✅ Load GeoFabrik artifact to db   1m 58s
   ◷ Initialize job   2s
   ⊘ Pre-job: Get Key Vault Secrets   <1s
   ✅ Checkout MapioCymruOGCStack...   1s
   ✅ Get Key Vault Secrets   <1s
   ✅ Make artifact download directory   3s
   ✅ Download geofabrik artifact   15s
   ✅ Download imposm binary artifact   10s
   ✅ list contents of new directory   1s
   ✅ run imposm import   1m 21s
   ✅ Post-job: Checkout MapioCymru...   <1s
   ◷ Finalize Job   <1s

∨ ✅ Load MapioCymru artifact to...   3m 29s
   ◷ Initialize job   2s
   ⊘ Pre-job: Get Key Vault Secrets   <1s
   ✅ Checkout MapioCymruOGCSta...   1s
   ✅ Get Key Vault Secrets   <1s
   ✅ Make mapiocymru artifact dow...   3s
   ✅ Download mapiocymru artifact   13s
   ✅ Check pg_restore is installed   1s
   ✅ Run pg_restore to load ma...   3m 3s
   ✅ Post-job: Checkout MapioCy...   <1s
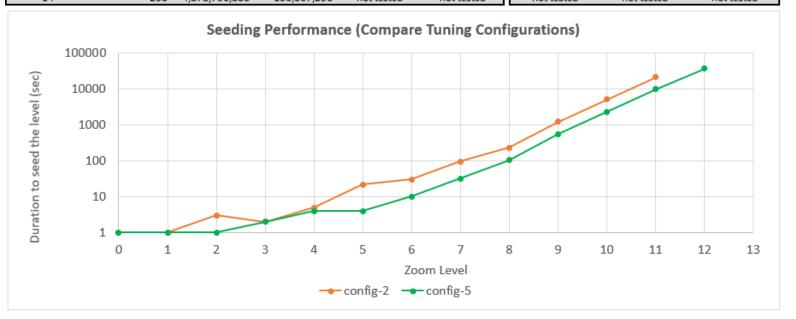   ◷ Finalize Job   <1s

Imposm importer loads data to postgres in imposm schema format

GeoFabrik PBF file → ETL A → Postgres db (imposm schema) → Updated PG database with MapioCymru data (imposm schema)

MapioCymru PG dump file (osm2pgsql schema) → ETL B → Postgres db (osm2pgSQL schema) → ETL C

pg restore loads MapioCymru data to Postgres in osm2pgsql schema format

Custom SQL to generate a lookup of "osm_id"and "name" from MapioCymru. Lookup table then JOIN + UPDATE to the IMPOSM schema tables which is compatible with the Geoserver stack.

Tuning Example

Optimise extents and parallelise

Header allows traffic to be directed to different app service instances behind the "invisible" load balancer

```
ARRAffinity=<app-service-instance-id>
```

```
# Post request to reseed cache for a specific AoI bounding box to app service instance-a (note need to change co
curl \
  --header "Content-type: application/json" \
  --cookie ARRAffinity=<app-service-instance-id-A> \
  --data '{"seedRequest":{"name":"osm:osm","bounds":{"coords":{"double":["168617","164433","356103","261969"]}}}.
  --user <geoserver-admin-username>:<geoserver-admin-password> \
                                           /geoserver/gwc/rest/seed/osm:osm.json

# Post request to reseed cache for a specific zoom level and specific AoI bounding box to app service instance-b
curl \
  --header "Content-type: application/json" \
  --cookie ARRAffinity=<app-service-instance-id-B> \
  --data '{"seedRequest":{"name":"osm:osm","bounds":{"coords":{"double":["210341","261969","355019","399061"]}}}.
  --user <geoserver-admin-username>:<geoserver-admin-password> \
                                           /geoserver/gwc/rest/seed/osm:osm.json
```

Optimise extents and parallelise

Performance

new gridset z level	new gridset Scale	new gridset total tiles within gridset bnds	new gridset blobs per zoom within AoI	new gridset config 2 per-zoom duration (sec)	new gridset config 2 per-zoom duration (hrs)	new gridset blobs per zoom within AoI	new gridset config 5 per-zoom duration (sec)	new gridset config 5 per-zoom duration (hrs)
0	10,000,000	4	4	1	0.00028	4	1	0.00028
1	5,000,000	16	16	1	0.00028	16	1	0.00028
2	2,000,000	80	32	3	0.00083	32	1	0.00028
3	1,000,000	320	64	2	0.00056	64	2	0.00056
4	500,000	1,248	96	5	0.00139	80	4	0.00111
5	250,000	4,928	256	22	0.00611	224	4	0.00111
6	100,000	30,528	1,280	30	0.00833	944	10	0.00278
7	50,000	122,112	4,320	97	0.02694	3,376	32	0.00889
8	25,000	488,448	16,240	229	0.06361	12,864	104	0.02889
9	10,000	3,050,880	100,224	1246	0.34611	76,688	562	0.15611
10	5,000	12,199,680	395,824	5248	1.45778	302,400	2325	0.64583
11	2,500	48,791,040	1,568,640	21603	6.00083	1,196,850	9931	2.75861
12	1,250	195,164,160	6,252,544	not tested	not tested	4,770,672	37504	10.41778
13	500	1,219,699,200	39,055,680	not tested	not tested	not tested	not tested	not tested
14	250	4,878,796,800	156,067,296	not tested	not tested	not tested	not tested	not tested

12 levels = 3.6 hrs

13 levels = 14 hrs
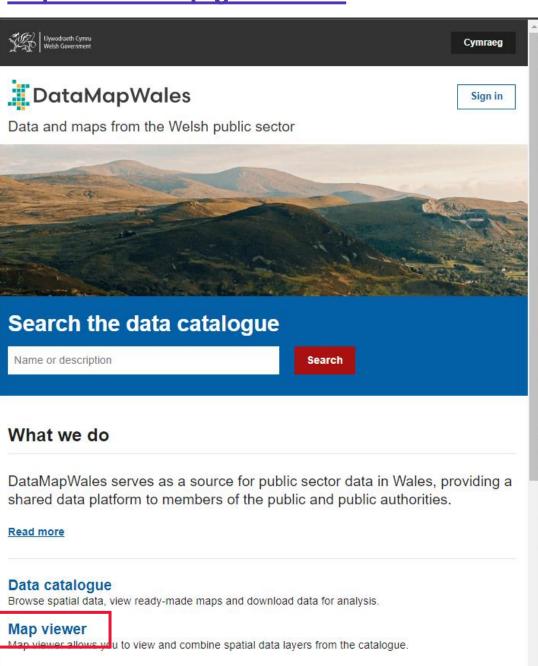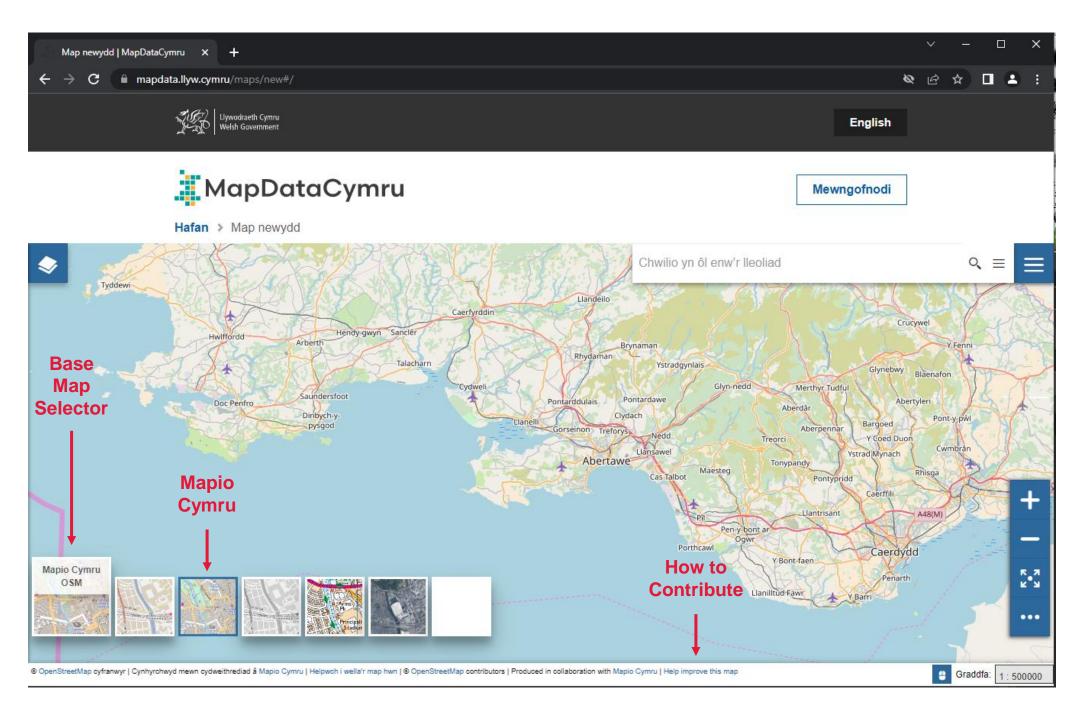
Seeding Performance (Compare Tuning Configurations)

Duration to seed the level (sec) vs Zoom Level — config-2, config-5

Go Live

Base Map Selector

Mapio Cymru

How to Contribute

Public

Thanks for listening!

Acknowledgements:
Welsh Government and Data Map Wales Team
Mapio Cymru team and Carl Morris
CGI team and James Fry
The FOSS4G community, particularly GeoSolutions
All past and future contributors to OSM

Internet-connection, question-mark, database, servers-group, cancel-48 icons and by icons8

Sam Franklin
sam.franklin@cgi.com
github.com/samfranklin