

Putting Cambridge GIS Data on GitHub

Spring NEARC - UMass Amherst
May 13, 2014

Sean Sweeney
City of Cambridge
@hiker4k

Open data is all the rage at all levels of government these days. There are a number of ways to provide GIS Open Data online, each one targeted to a specific audience. One of the ways Cambridge GIS is sharing their data is through the social coding site GitHub.

For developers, GitHub is a familiar place where they already share code. Commercial companies such as Esri have embraced GitHub as a platform for their open source software offerings. To facilitate the sharing of data as well as code, GitHub has built in a number data management and visualization features, including map-based visualizations for GIS data.

I will take you on a tour of our GitHub data repository and show you how we created and maintain it.



Why open data?

1. Citizens demand it.
2. City management wants to provide it.
3. Civic engagement organizations request it (e.g. Code for America).
4. We have lots of students who will seek it out.
5. We want people to have access to and use our data.

We work hard to create and maintain these data sets and the more they get out there the more potential value they can provide to the community.



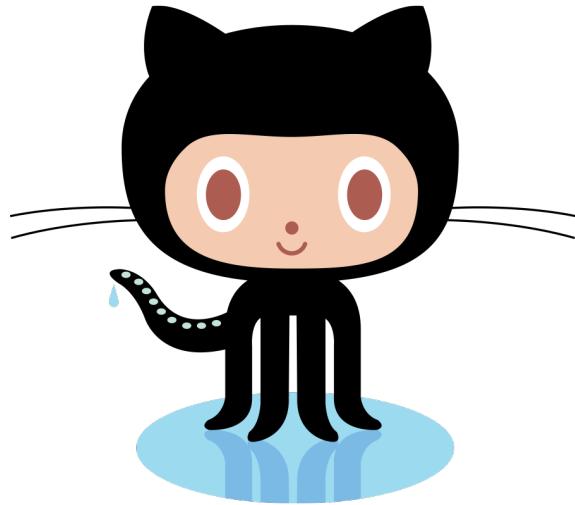
Why GeoJSON?

Why GeoJSON?

It's webby!

We have our data available for download in Shapefile and File Geodatabase format as well, but GeoJSON is a format that can be understood and easily consumed by web developers whether or not they are GIS programmers.

Why GitHub?



Why GitHub?

“GitHub is the best place to share code with friends, co-workers, classmates, and complete strangers. Over three million people use GitHub to build amazing things together.”

- <https://github.com/about>

GitHub is familiar to developers and provides a number of advantages for storing open data.

- Any data
 - Hosted
 - Downloadable
 - Change tracking
 - API access
- GIS data
 - Visualize
 - Format supported by all major map clients

The screenshot shows a GitHub showcase page titled "Open source organizations". At the top, it displays statistics: 14 Repositories, 4 Languages, and was last updated 2 months ago. Below this is a green header bar with the title. The main content area lists several organizations with their GitHub URLs and brief descriptions:

- adobe/adobe.github.com**: Adobe central hub for open source. A site for the CFPB to share and discuss its technology work with the world.
- cfpb/cfpb.github.io**: A site for the CFPB to share and discuss its technology work with the world.
- Netflix/netflix.github.com**: A listing of open source efforts at Netflix on GitHub.
- Esri/esri.github.com**: Esri on GitHub.
- square/square.github.io**: A simple, static portal which outlines our open source offerings.
- twitter/twitter.github.com**: A listing of open source efforts at Twitter on GitHub.

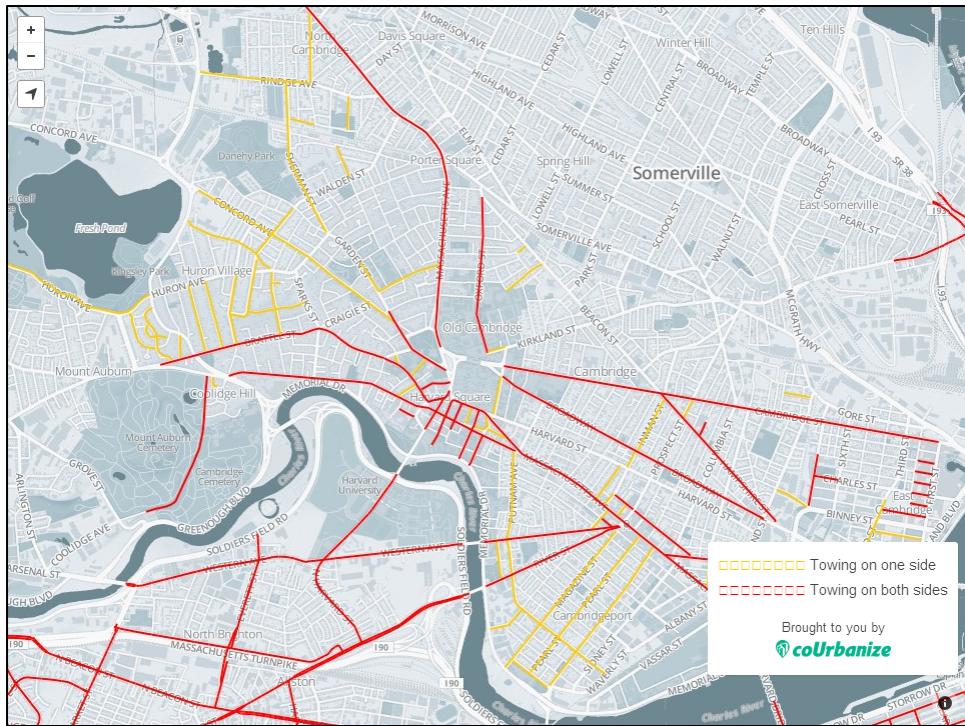
On the right side, there are sections for "Related Showcases" (Projects that power GitHub, Government apps) and "New Showcases" (Emoji, Projects with great wikis). A search bar and a "View all" link are also present.

Who is using GitHub?

- [GitHub open source organizations](#)
- For data:
 - [Chicago](#)
 - [Philadelphia](#)
 - [Bloomington, IN](#)

Tour





Tour: Usage examples - [Will they tow me?](#)

Everybody in Cambridge hates parking tickets.
Where can you park your car and not get a ticket?
Find out by browsing this street-cleaning map below.

Enter a location Search Maps

307 Broadway, Cambridge, MA 02139, USA
Street cleaning: APR 7, MAY 5, JUNE 2, JULY 7, AUG 4, 29, OCT 6, NOV 3, DEC 1, APR 1, MAY 6, JUNE 3, JULY 1, AUG 5, SEP 2, OCT 7, NOV 4, DEC 2
[Setup email alerts](#) or [Add to Calendar](#)

The city of Cambridge, MA has street cleaning running from April 1st till December 31 each year.

Online

[instantcheckmate.com](#)

Search Criminals In Massachusetts.
Check Their Records Online Today.

Switch To A New Fiat 500L

Traffic Court Tickets

NYC Parking

Print Free Buffet Coupons

Coupon Savings At Target

Free Day of Parking

Find Military Records

Tour: Usage example - [Cambridge Street Cleaning](#)



Cambridge GIS

City of Cambridge, MA

[Cambridge GIS GitHub Home](#)

[View My GitHub Profile](#)

Open Data on GitHub

The City of Cambridge is committed to building repositories for data to be used by developers and for public consumption. Cambridge GIS has posted the majority of the data sets on this official City of Cambridge site. Larger size files which do not lend themselves to this site (such as elevation data, raster, and LiDAR) will soon be available for download at www.cambridgema.gov/GIS and also available on DVD or by stopping by our offices (by appointment) at 831 Massachusetts Avenue.

GeoJSON

GitHub recognizes our GeoJSON files and creates interactive maps using Mapbox for most files. You can also view or download the raw GeoJSON using the 'Raw' button.

Report Issues and Update Data

If you notice a problem with the data, you can report an issue on this site using the [New Issue form](#).

You can also click 'Edit' on any GeoJSON file to make a fork/copy on your own account, make changes, and suggest that they get adopted here.

About the City of Cambridge GIS Data

All of our data is in Massachusetts State Plane Coordinate System NAD 83, feet, except those files converted to GeoJSON which are in WGS84. We are constantly upgrading and improving our data layers and hopefully you will find this data useful and accurate. Please notify Cambridge GIS if you see any issues with the data by contacting us at CambridgeGIS@cambridgema.gov.

City of Cambridge, MA GIS Data Sets Available on GitHub

Tour: [GitHub Pages](#)

The screenshot shows the GitHub organization page for 'City of Cambridge GIS'. At the top, there's a circular logo with 'GIS' in the center and 'CITY OF CAMBRIDGE MASSACHUSETTS' around it. The main heading is 'City of Cambridge GIS'. Below the logo, it says 'Cambridge, MA' and provides a link to their website: <http://www.cambridgemass.gov>. There are search and filter buttons, and a 'New repository' button. On the right, there are sections for 'Members' (with Sean Sweeney and Sean Farber listed) and 'Teams' (with Team AddressDashboard and Team ags-tools). The organization has four repositories listed:

- cambridgegis_data**: Shell, 33 stars, 17 forks. Description: City of Cambridge GIS Data. Updated 5 days ago.
- cambridgegis_data_dpw**: Shell, 0 stars, 0 forks. Description: Cambridge GIS data for the DPW category only. Updated 5 days ago.
- cambridgegis_data_infra**: Shell, 0 stars, 0 forks. Description: Cambridge GIS data for the Infrastructure category only. Updated 5 days ago.
- cambridgegis.github.io**: CSS, 0 stars, 0 forks. Description: GitHub Pages for Cambridge GIS. Updated 16 days ago.

Tour: Organization

Organizations:

- Group many users together
- Are administered by users
- Can own repositories
- Contains teams (if you want)

Example: [The White House](#)

- [Ben Balter, GitHub](#)

The screenshot shows the GitHub repository page for `cambridgegis / cambridgegis_data`. The main content area displays a table of 40 commits on the `master` branch. Each commit details an update to a specific GIS layer, such as Address, Assessing, Basemap, Boundary, CDD, DPW, Demographics, Elections, Health, Historical, Hydro, Infra, Landmark, Public_Safety, Recreation, and Traffic. The commits are timestamped from 6 days ago to 9 months ago. The right sidebar includes links to Code, Issues, Pull Requests, Wiki, Pulse, Graphs, Network, and Settings, along with clone and download options.

Layer	Description	Time Ago
Address	Data updated for the following layers:	a month ago
Assessing	Updated README files for parcel layers.	3 months ago
Basemap	Data updated for the following layers:	a month ago
Boundary	Added remaining files upload.	9 months ago
CDD	Data updated for the following layers:	3 months ago
DPW	Data updated for the following layers:	6 days ago
Demographics	Added remaining files upload.	9 months ago
Elections	Updated Congressional Districts to reflect election of Katherine Clar...	5 months ago
Health	Added remaining files upload.	9 months ago
Historical	Added remaining files upload.	9 months ago
Hydro	Added remaining files upload.	9 months ago
Infra	Data updated for the following layers:	6 days ago
Landmark	Data updated for the following layers:	a month ago
Public_Safety	Added remaining files upload.	9 months ago
Recreation	Data updated for the following layers:	a month ago
Traffic	Added the large layers back in GeoJSON format.	8 months ago

Tour: [Repository](#)

GitHub Repositories:

- Most basic element of GitHub
 - Contains files
 - Stores revision history
 - Displays documentation
 - Can have multiple collaborators
- [Ben Balter, GitHub](#)

The `cambridgegis_data` repository contains all of our data in one big download.

Cambridge GIS data for the DPW category only. — Edit

14 commits 1 branch 0 releases 1 contributor

branch: master cambridgegis_data_dpw

Data updated for the following layers: ...

Litter_Buckets	Data updated for the following layers:	2 months ago
Pedestrian_Ramps	Data updated for the following layers:	6 days ago
Snow_Emergency_Tow_Rou...	Added Snow Emergency Tow Routes layer with the following disclaimer:	4 months ago
Street_Sweeping_Districts	Added Street Sweeping.	17 days ago
Street_Sweeping_Roads	Added Street Sweeping.	17 days ago
Trees	Data updated for the following layers:	6 days ago
README.md	Update README.md	7 months ago

README.md

Cambridge GIS Data - DPW

This is the Cambridge GIS data for the DPW category only. For the complete Cambridge GIS dataset see the [cambridgegis_data](#) repository. The [README](#) in that repository contains more details about the Cambridge GIS open data.

Code Issues Pull Requests Wiki Pulse Graphs Network Settings

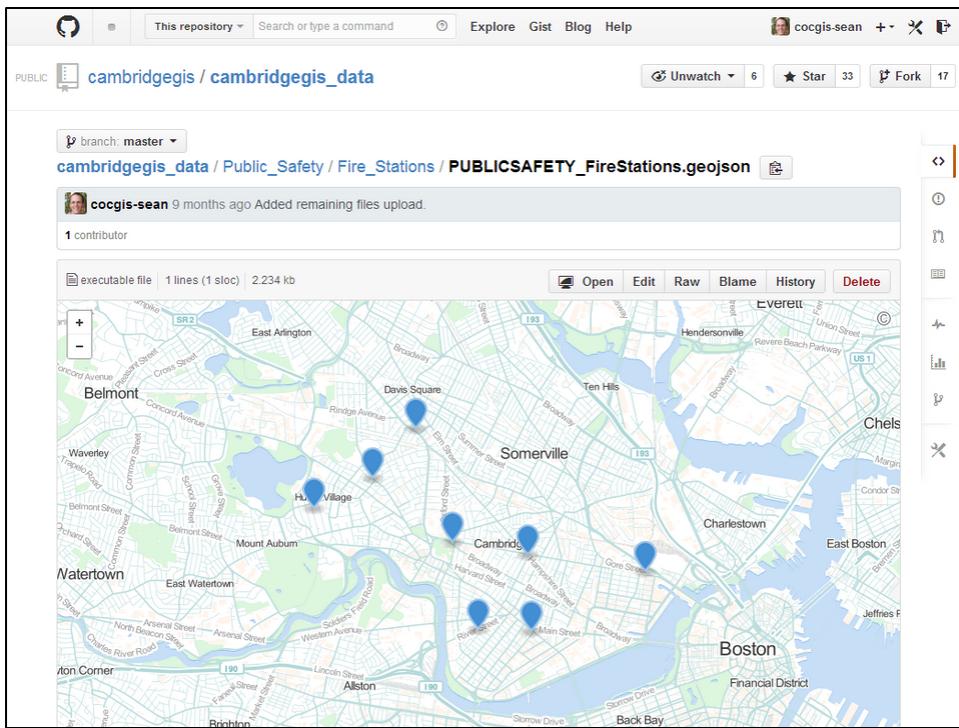
HTTPS clone URL https://github.com/cocgis-sean/cambridgegis_data_dpw

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

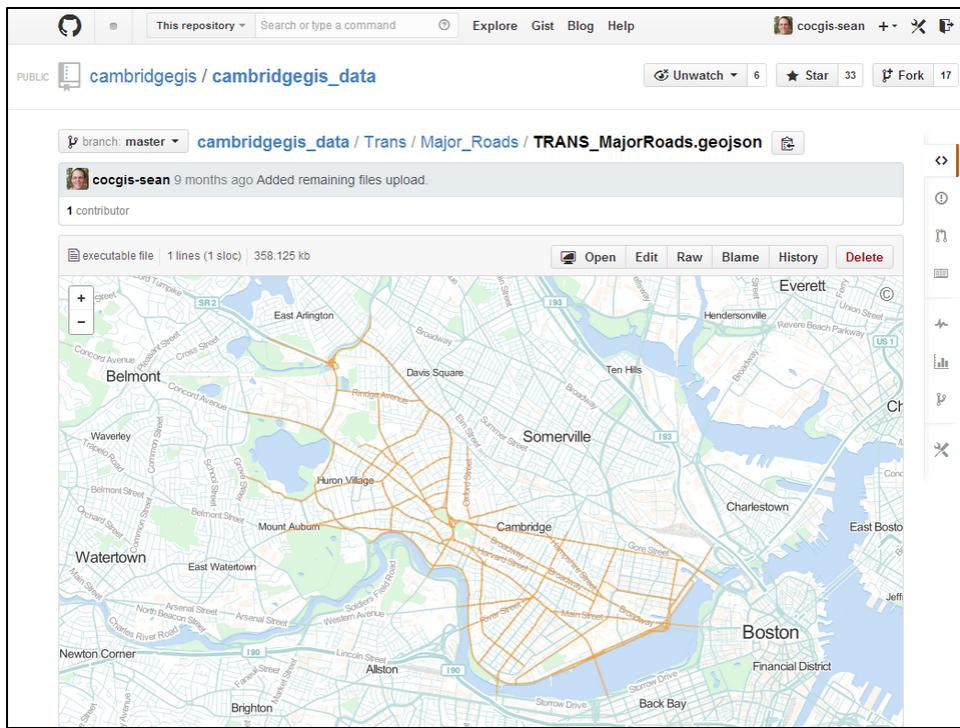
Clone in Desktop Download ZIP

Tour: Sub-repo

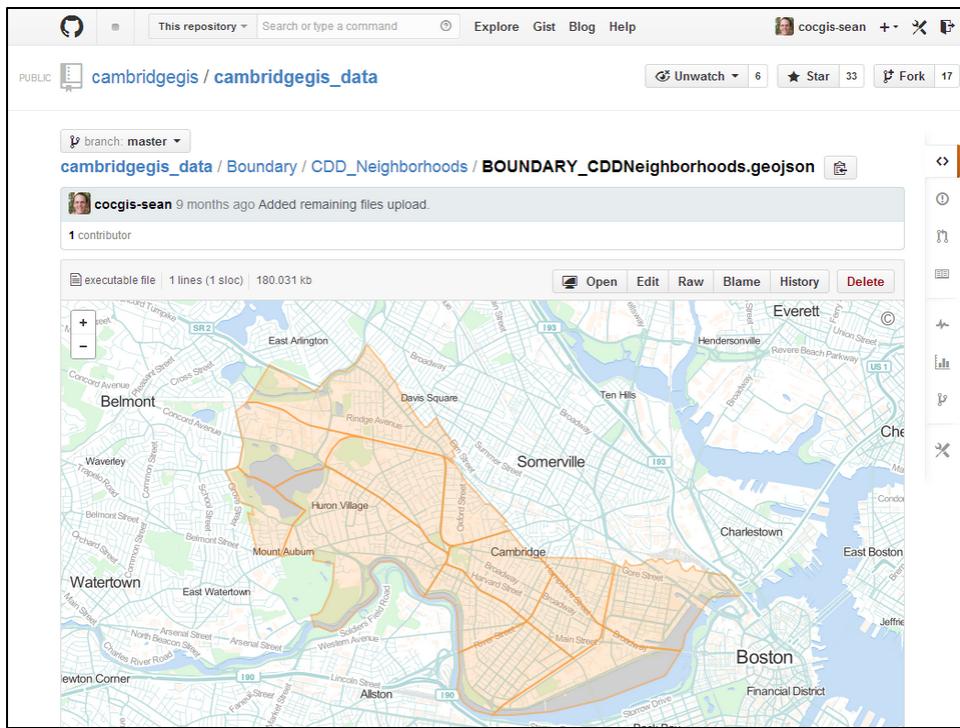
We have sub-repositories by category so that Smaller chunks of data can be cloned.



Tour: [Fire Station Points](#)



Tour: [Major Roads Lines](#)



Tour: [Neighborhood Polygons](#)

The screenshot shows a GitHub repository page for 'cambridgegis / cambridgegis_data'. The repository has 33 stars and 18 forks. A commit by 'cogis-sean' from 9 months ago added the file 'BOUNDARY_CDDNeighborhoods.xml'. The file is an executable XML file with 207 lines and 6.224 kb size. The XML code describes the boundaries of Cambridge's neighborhoods, including the title 'BOUNDARY_CDDNeighborhoods', the purpose ('Created for the establishment of neighborhood planning groups, administration of Community Development Block'), and the geographic coordinates of the bounding box.

```
1 <metadata>
2   <idinfo>
3     <citation>
4       <citeinfo>
5         <title>BOUNDARY_CDDNeighborhoods</title>
6         <geoform>vector digital data</geoform>
7       </citeinfo>
8     </citation>
9   <descript>
10    <abstract>Created for the establishment of neighborhood planning groups, administration of Community Development Block</abstract>
11    <purpose>This polygon layer contains the boundaries of the City of Cambridge's thirteen neighborhoods.</purpose>
12  </descript>
13  <spdom>
14    <bounding>
15      <westbc>-71.160581</westbc>
16      <eastbc>-71.063810</eastbc>
17      <northbc>42.352358</northbc>
18      <southbc>42.352358</southbc>
19    </bounding>
20  </spdom>
21  <keywords>
22    <theme>
23      <themekt>None</themekt>
24      <themekey>City of Cambridge</themekey>
25      <themekey>CDD</themekey>
26      <themekey>Boundary</themekey>
27      <themekey>Neighborhoods</themekey>
28    </theme>
29  </keywords>
30  <accconst>None</accconst>
31  <useconst>None</useconst>
```

Tour: [Metadata](#)

The screenshot shows the GitHub Issues page for the repository 'cambridgegis/cambridgegis_data'. The interface includes a header with navigation links like 'Explore', 'Gist', 'Blog', and 'Help'. On the right, there are user stats: 'cocgis-sean' (6 issues), 'Star 33', and 'Fork'. The main area displays a list of issues with columns for title, number, and status. A sidebar on the left provides filtering options for 'Everyone's Issues' (5) and 'Labels' (including 'bug', 'duplicate', 'enhancement', 'invalid', 'question', and 'wontfix'). A 'New label' input field is also present.

Issue	Number
Storm Drainages	#7
Snow emergency route data	#5
Link to new issue form	#4
Post the larger geojson files	#3
Split the repo up	#2

Tour: [Issues](#)

Issues:

- Suggested improvements, tasks, action items
- Created by users, closed by admins
- Contain their own discussion forum
- Grouped by tags and milestones
- “Tickets” in traditional PM speak

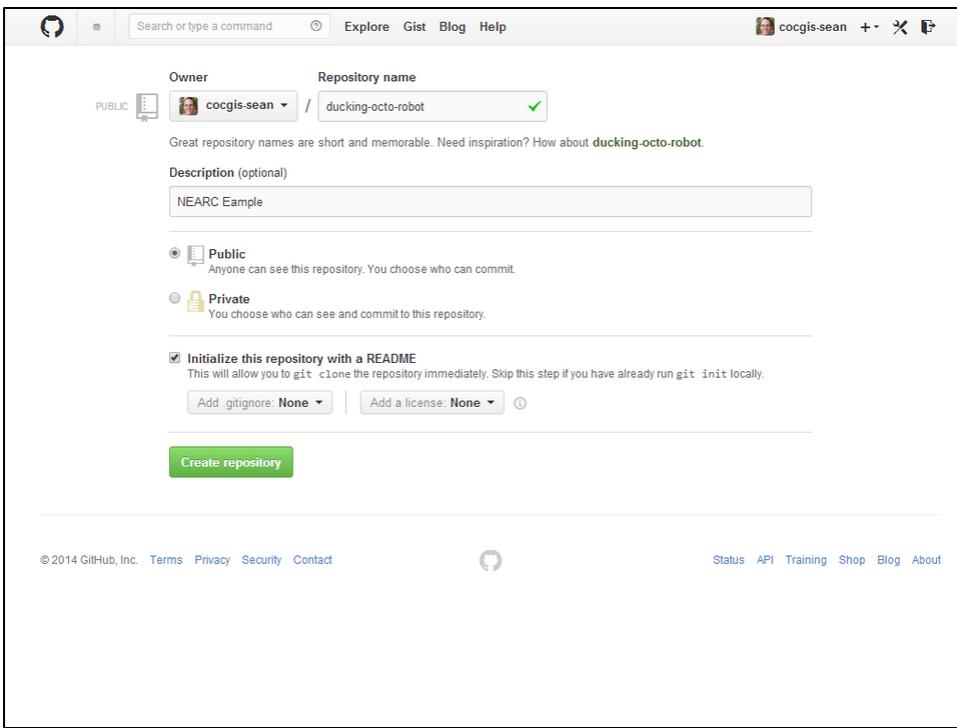
- [Ben Balter, GitHub](#)



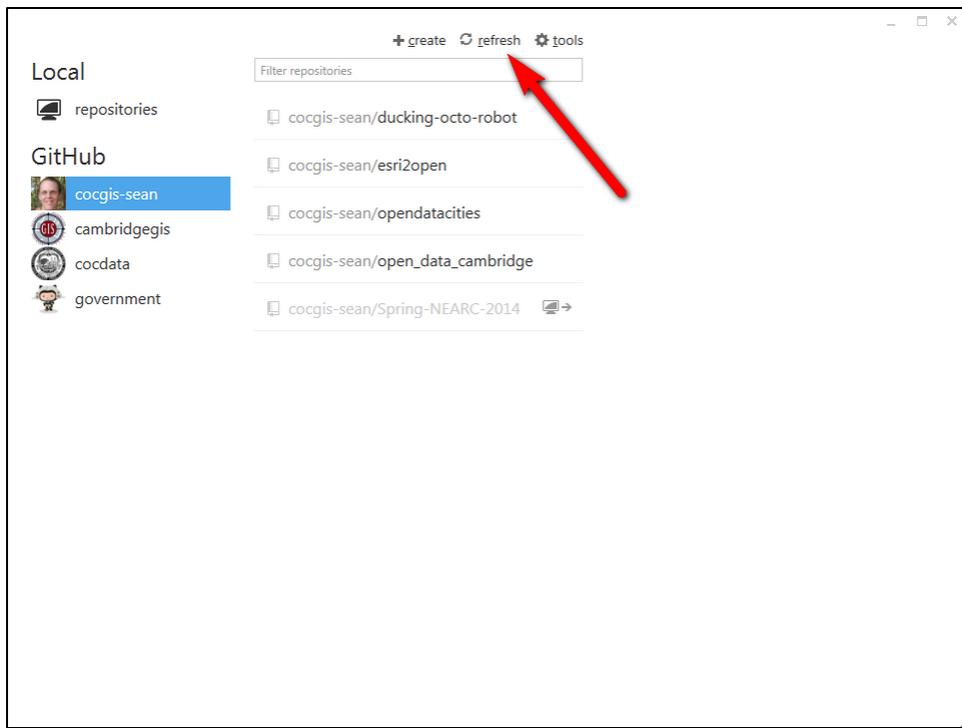
Tour: [GitHub for Windows](#)

The screenshot shows a GitHub organization page for 'City of Cambridge GIS'. The top navigation bar includes 'Search or type a command', 'Explore', 'Gist', 'Blog', and 'Help'. The user 'cocgis-sean' is logged in. A dropdown menu on the right offers options like 'New repository', 'New organization', 'New member', 'New team', and 'New repository'. The main content area displays four repositories: 'cambridgegis_data' (Shell, 33 stars, 17 forks), 'cambridgegis_data_dpw' (Cambridge GIS Data for the DPW category only, 0 stars, 0 forks), 'cambridgegis_data_infra' (Cambridge GIS data for the Infrastructure category only, 0 stars, 0 forks), and 'cambridgegis.github.io' (GitHub Pages for Cambridge GIS, CSS, 0 stars, 0 forks). On the right sidebar, sections for 'Members' (cocgis, Sean S, sfarber) and 'Teams' (AddressDashboard, ags-tools) are shown, along with a 'Create a team' button.

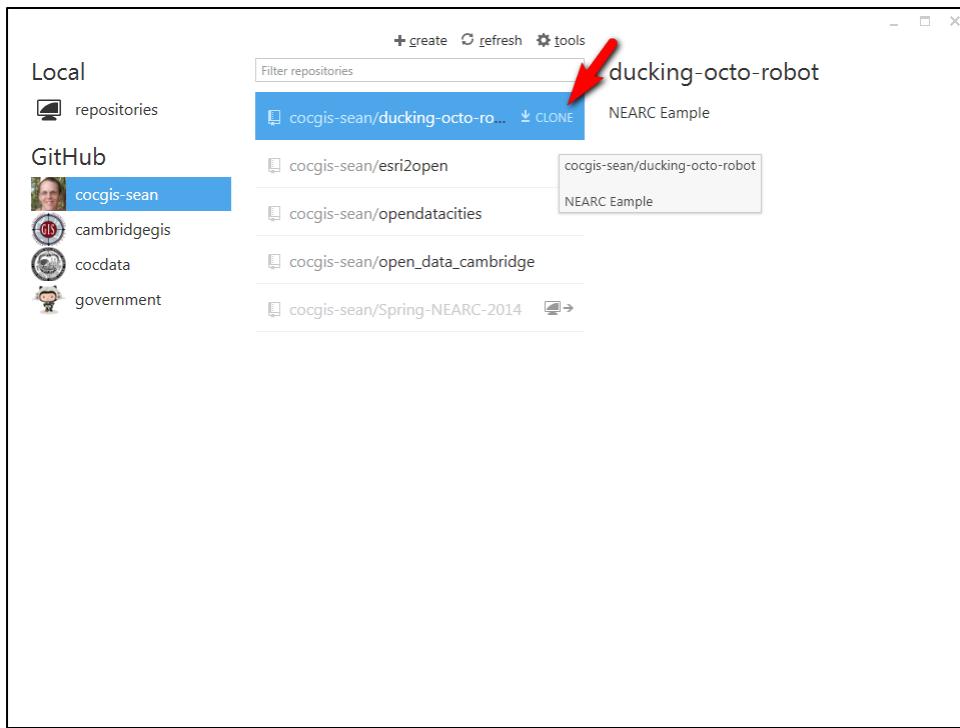
Tour: Creating the repo



Tour: Creating the repo



Tour: Syncing down



Tour: Cloning the repo

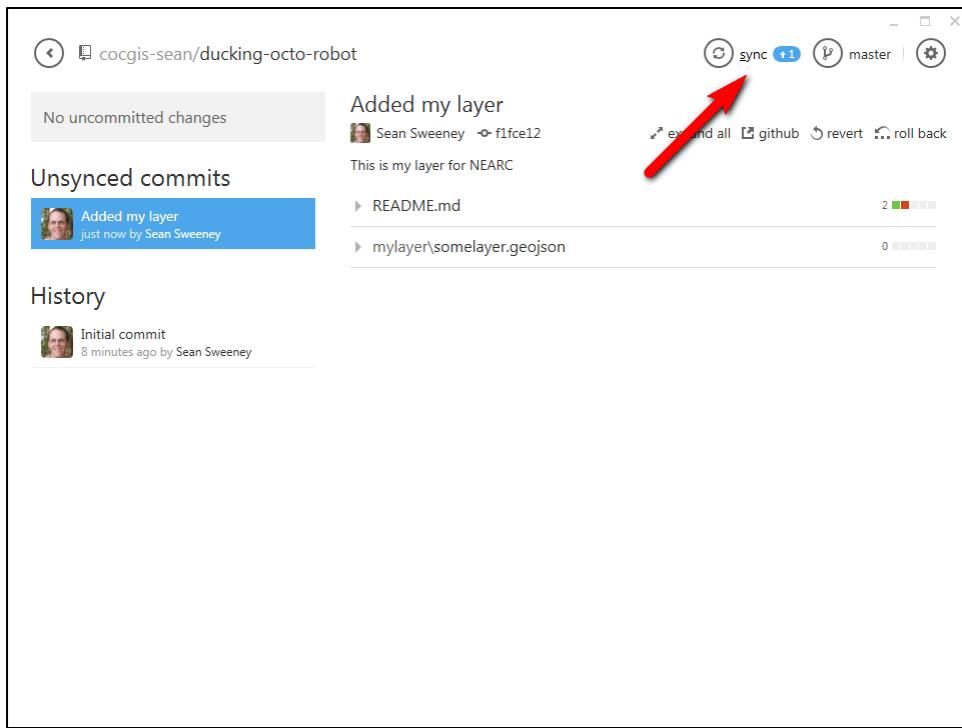
The screenshot shows the GitHub desktop application interface. At the top, it displays the repository 'cocgis-sean/ducking-octo-robot'. In the top right corner, there are status icons for 'sync' (with 1 update), 'master', and settings. Below the header, there's a section for 'Uncommitted changes' with a 'hide' button. A message box says 'Made some changes' with a text area containing 'Here is what I did:' followed by a redacted content area. A large red arrow points from this area to the 'Commit to master' button, which is highlighted with a grey background and contains the text 'Commit to master' and '1 file to be committed'. To the left of this, under 'Unsynced commits', is a commit from 'Sean Sweeney' adding a layer. Below this is a 'History' section showing an 'Initial commit' by Sean Sweeney 4 hours ago. On the right, there's a 'Files to commit' section for 'README.md' with a diff view. The diff shows changes from 'ducking-octo-robot' to 'Ducking Octo Robot'. The bottom right of the interface has a 'collapse all' button.

Tour: Viewing changes and committing locally commit.

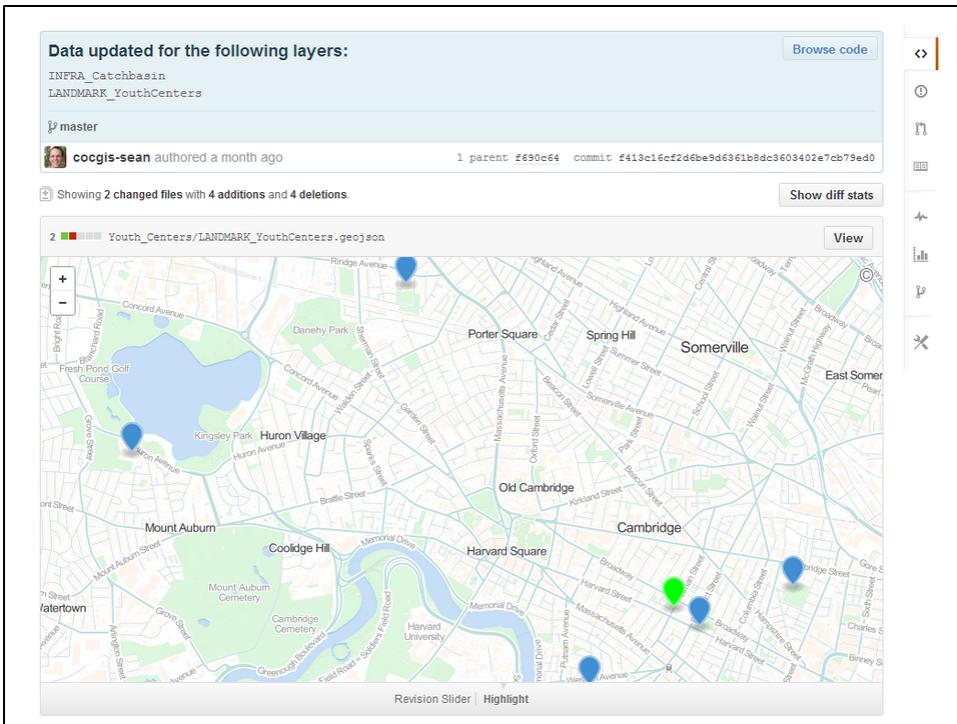
Commits:

- Individual change to a file(s) in Repository
- Uniquely identifiable
- Who made the change
- When was it made
- Can be compared to other commits

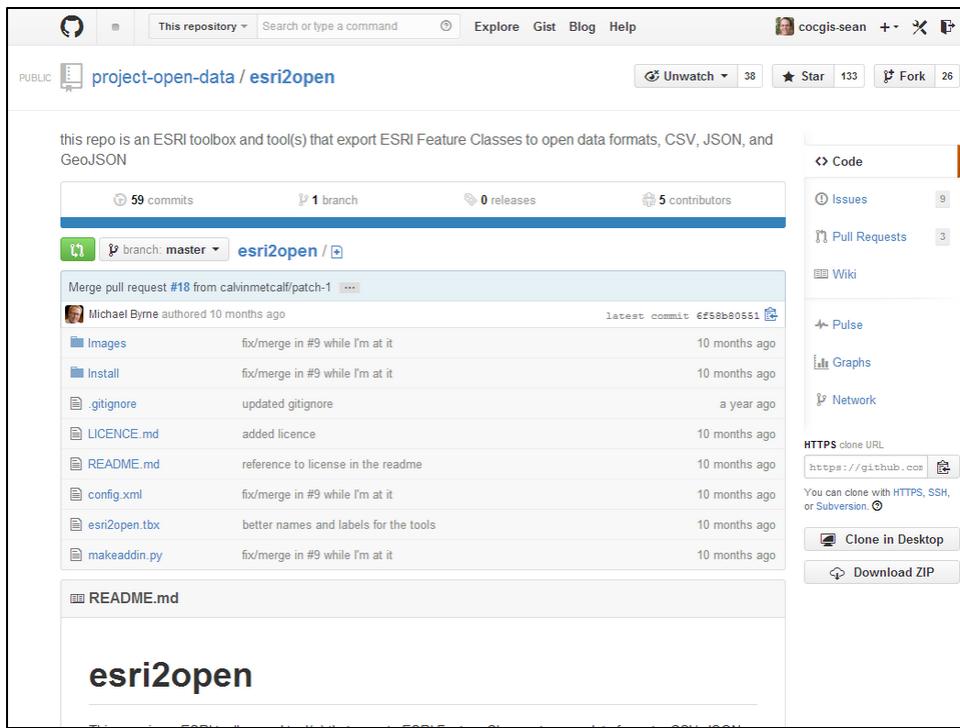
- [Ben Balter, GitHub](#)



Tour: Pushing updates



Tour: [Showing Diffs](#)



Tour: [Esri2Open](#)

- Michael Byrne, Consumer Finance Protection Bureau
- Calvin Metcalf, AppGeo (formerly MassDOT)

```
for db, ds, fc, dest_name, dest_path) in layers:  
    # Change workspaces as needed. Not sure of the cost of this so I sorted the  
    # input list by db a  
    if not db in arcpy.env.workspace:  
        print "Changing to db: " + db  
        arcpy.env.workspace = workspace_path + db + '.sde'  
  
    if ds != "":  
        source = db + "." + schema + "." + ds + "/" + db + "." + schema + "." + fc  
    else:  
        source = db + "." + schema + "." + fc  
  
    dest = dest_path + "\\\" + dest_name  
  
    if not arcpy.Exists(source):  
        print "ERROR: " + source + " Not Found"  
    else:  
        print "Exporting: " + source  
        print "To: " + dest  
        arcpy.gp.esri2open_esri2open(source, dest + ".geojson", "Default")  
        arcpy.ExportMetadata_conversion(source, translator, dest + ".xml")  
        print "--"
```

Tour: ArcPy export/update script



Geographic Information System

City of Cambridge, MA

Contact GIS Cambridge Home Page

Text Size: A A A

Search



INTERACTIVE MAPS

MYCAMBRIDGE

MAP GALLERY

GIS DATA

MOBILE MAPS

ABOUT GIS

FunLink

GIS DATA DICTIONARY

Address

Assessing

Basemap

Boundary

CDD

Demographics

DPW

Elections

Health

Historical

Hydro

Images

Infra

Landmark

Public Safety

Recreation

Topography

Traffic

GIS > GIS Data Dictionary

GIS Data Dictionary

We are in the process of creating an all new Cambridge GIS Data Dictionary. These pages are about a year old, but do not include data from our new April 14, 2010 flyover. Please keep checking back for updates.

The GIS Database Dictionary is an online guide for our GIS user community. Each page has detailed information on the individual GIS data layers that are stored on the City of Cambridge GIS system.

The dictionary contains information on how and why a GIS layer was created, coding of attributes, our procedure for maintaining each layer, departments which contribute to the development of each layer, history, and the intended use of the data.

- Address
- Assessing
- Basemap
- Boundary
- CDD
- Demographics
- DPW
- Elections
- Health
- Historical
- Hydro
- Images
- Infra
- Landmark
- Public Safety
- Recreation
- Topography
- Traffic
- Trans

Coming attractions for open data

New Cambridge GIS Data Dictionary: <http://www.cambridgema.gov/gis/gisdatadictionary>

ArcGIS Open Data

Connecting Citizens and Businesses
to Your Authoritative Data



Leverage Your Investment

Coming attractions for open data

ArcGIS Open Data .beta: <http://blogs.esri.com/esri/arcgis/2014/04/24/arcgis-open-data-beta/>

Slides & more info

<http://git.io/oM4v8A>

(search #nearc for @hiker4k on twitter)

ssweeney@cambridgema.gov