Overview

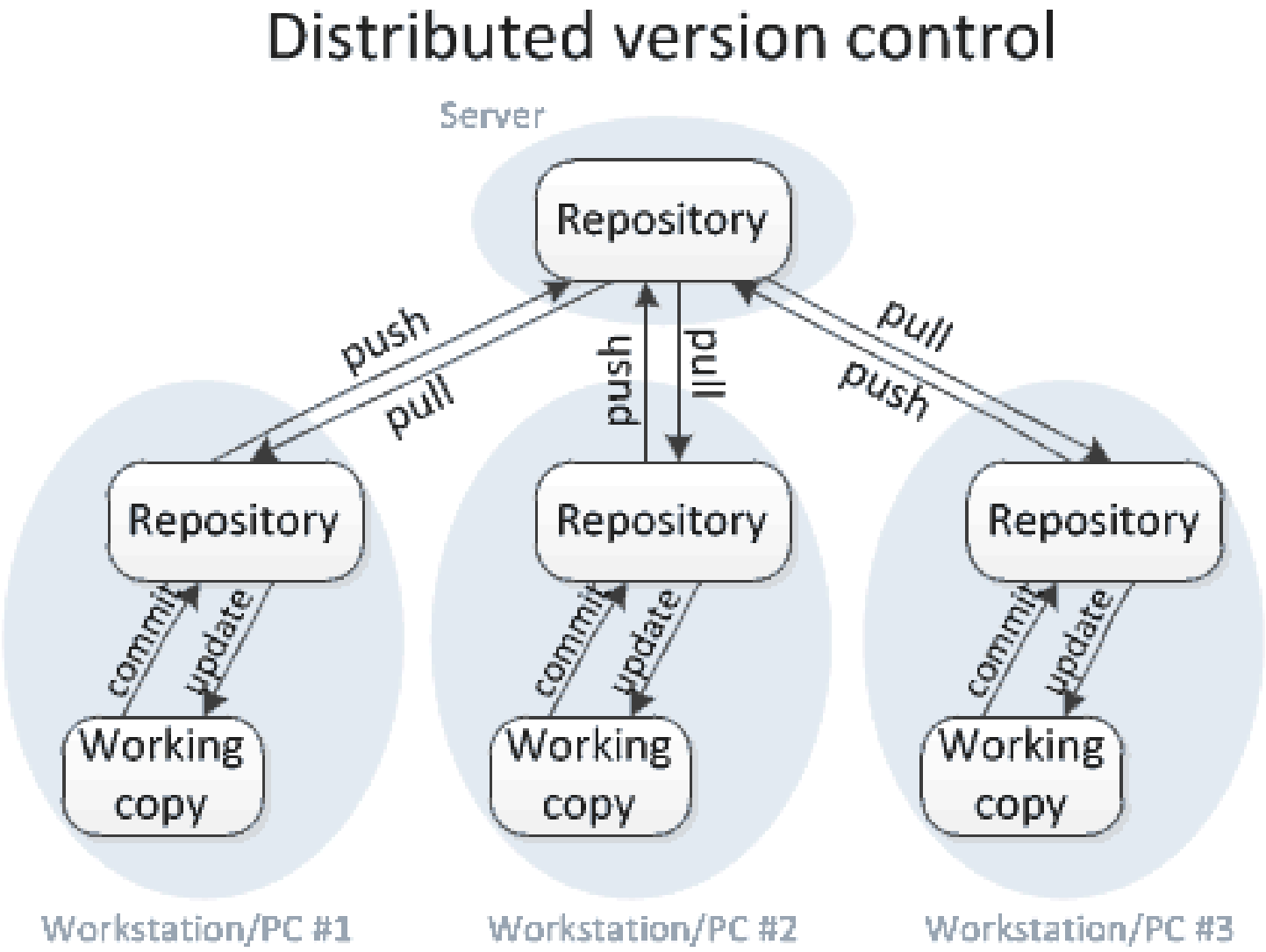
1. GitHub For Open GIS Data
2. ArcOpen
3. Publishing Open Data
4. Getting Started

- Offers features like issue tracking, analytics and social network capabilities
- Used by over 4 million people on a total of 10.2 million repositories
- Increasingly being used for not only hosting code but data, like at the City of Philadelphia

Quick explanation

- **Git**: an open source distributed version control system
- **Github.com**: Git as a Service and social network
- **Repository ("repo")**: a directory where a project (or data) lives
- **Commit**: a snapshot of work of the changes that you made to a repository
- **Pull request**: a request to a repository admin to accept your changes into their repo
- **`git clone`**: Making a copy of a remote repository to your machine
- **`git push`**: Committing your changes to the canonical copy of the repo on Github.com
- **`git pull`**: Bringing in a copy of the repo to your local machine

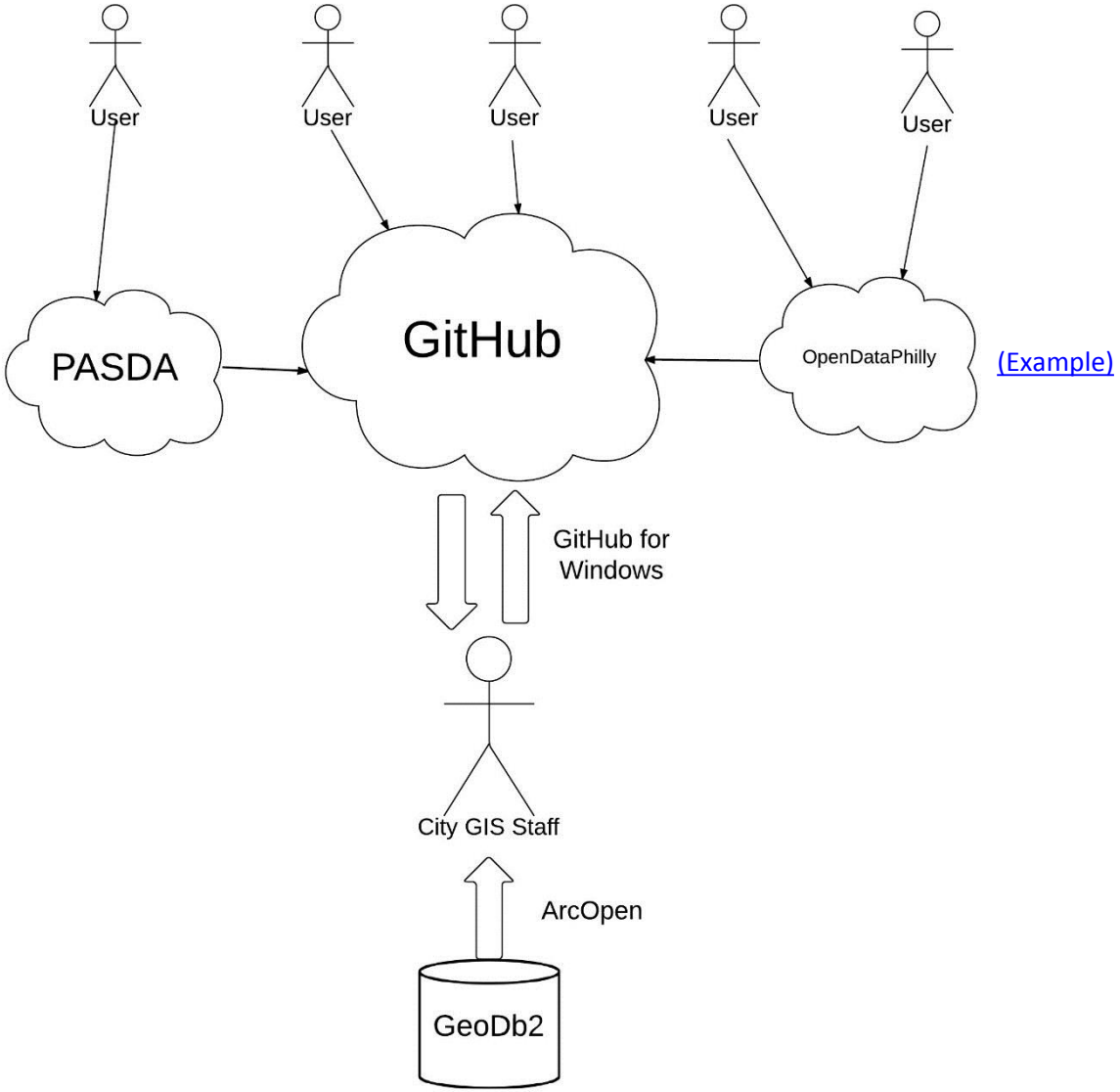
Distribution Version Control



Benefits of GitHub

- Gives GIS units control of when and how their data is updated to the general public
- Allows other GitHub users to ask questions and raise issues about the data - creating a dialog around it
- Displays a history of changes to the data that gives consumers notifications on updates
- Allows for collaboration between staff and outside stakeholders

Open Data Flow



Exporting Geodata…

Goals

- Provide an ArcGIS Toolbox tool that makes it easy to:
  - Publish data in a variety of formats (CSV, GeoJSON, KMZ, zipped and unzipped shapefiles)
  - Generate a README.md Markdown file based off of the metadata of the layer
  - Convert the dataset to WGS84 – lat/lng (4326), if necessary
- Can be used on data from GeoDb2 or anywhere else

GeoJSON

- Geographic features in JavaScript Object Notation (JSON)
- Primary use: Web mapping; the easiest format for visualization and analysis in most programming languages


```
{
    "type": "FeatureCollection",
    "features": [{
        "geometry": {
            "type": "Point",
            "coordinates": [-75.15351600153949, 40.06429504707634]
        },
        "type": "Feature",
        "id": 0,
        "properties": {
            "FMN": "Yes",
            "OPERATOR": "The Food Trust",
            "ADDRESS": "72nd Ave and Ogontz Ave, 19138",
            "NAME": "West Oak Lane",
            "SNA": "Yes"
        }
    },
    ...
    }]
}
```

KMZ

- Compressed Keyhole Markup Language, an XML notation
- Primary use: Google Earth and Google Maps

```
<?xml version="1.0" encoding="UTF-8"?>
   <kml xmlns="http://www.opengis.net/kml/2.2">
      <Document>
         <Placemark>
            <name>New York City</name>
            <description>New York City</description>
            <Point>
               <coordinates>-74.006393,40.714172,0</coordinates>
            </Point>
         </Placemark>
      </Document>
   </kml>
   ...
```

CSV

- Comma separated values
- Primary use: Excel and other spreadsheet-based tools; easy parsing and importing into a spatial database
- Only makes sense for point feature classes where lat and lngs are columns in the spreadsheet

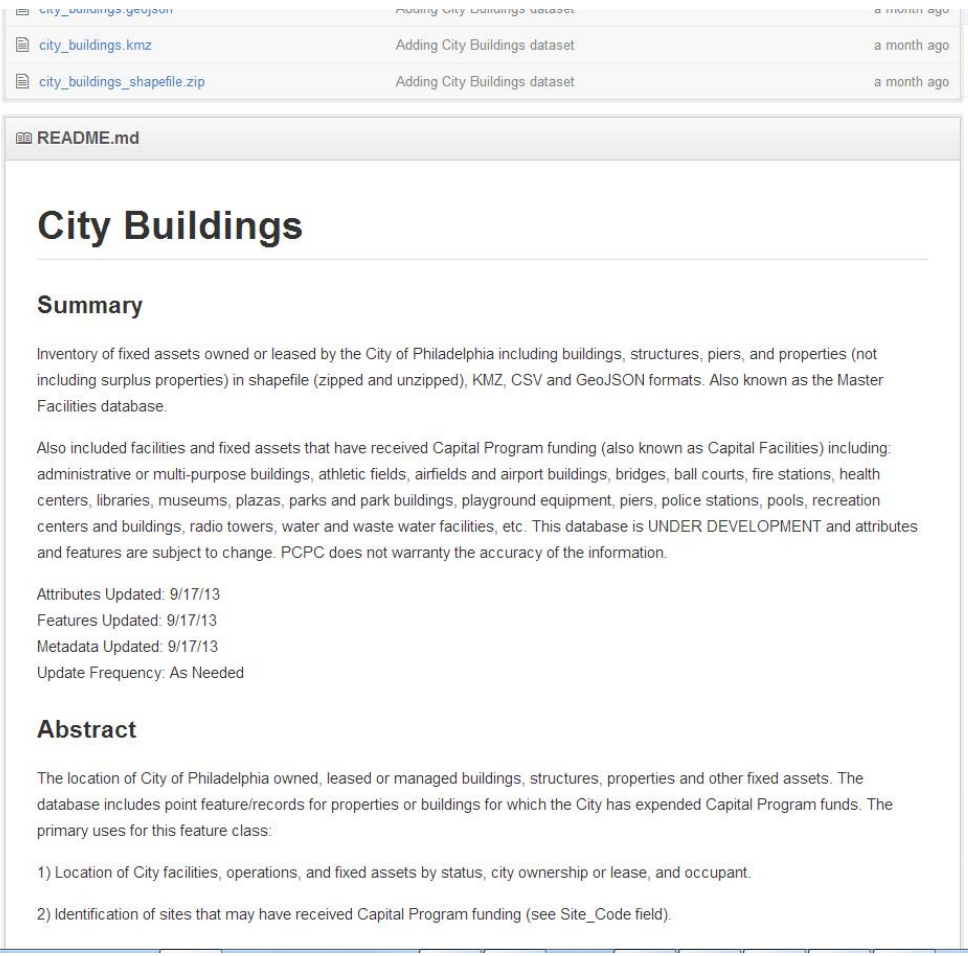
Shapefile

- Esri's geospatial vector format
- Primary use: Esri software and other desktop GIS suites
- Since a shapefile is made up of potentially seven files, it is also a good idea to provide a compressed (.zip) version of the shapefile
- Also, shapefile column names can be only 10 characters max. We should avoid column names getting truncated where possible.

Markdown

- A markup language for HTML
- Primary use: native rendering of text on github.com
- Metadata about a dataset should be saved in a README.md that will automatically be displayed on github.com

Markdown



city_buildings.kmz	Adding City Buildings dataset	a month ago
city_buildings_shapefile.zip	Adding City Buildings dataset	a month ago

README.md

City Buildings

Summary

Inventory of fixed assets owned or leased by the City of Philadelphia including buildings, structures, piers, and properties (not including surplus properties) in shapefile (zipped and unzipped), KMZ, CSV and GeoJSON formats. Also known as the Master Facilities database.

Also included facilities and fixed assets that have received Capital Program funding (also known as Capital Facilities) including: administrative or multi-purpose buildings, athletic fields, airfields and airport buildings, bridges, ball courts, fire stations, health centers, libraries, museums, plazas, parks and park buildings, playground equipment, piers, police stations, pools, recreation centers and buildings, radio towers, water and waste water facilities, etc. This database is UNDER DEVELOPMENT and attributes and features are subject to change. PCPC does not warranty the accuracy of the information.

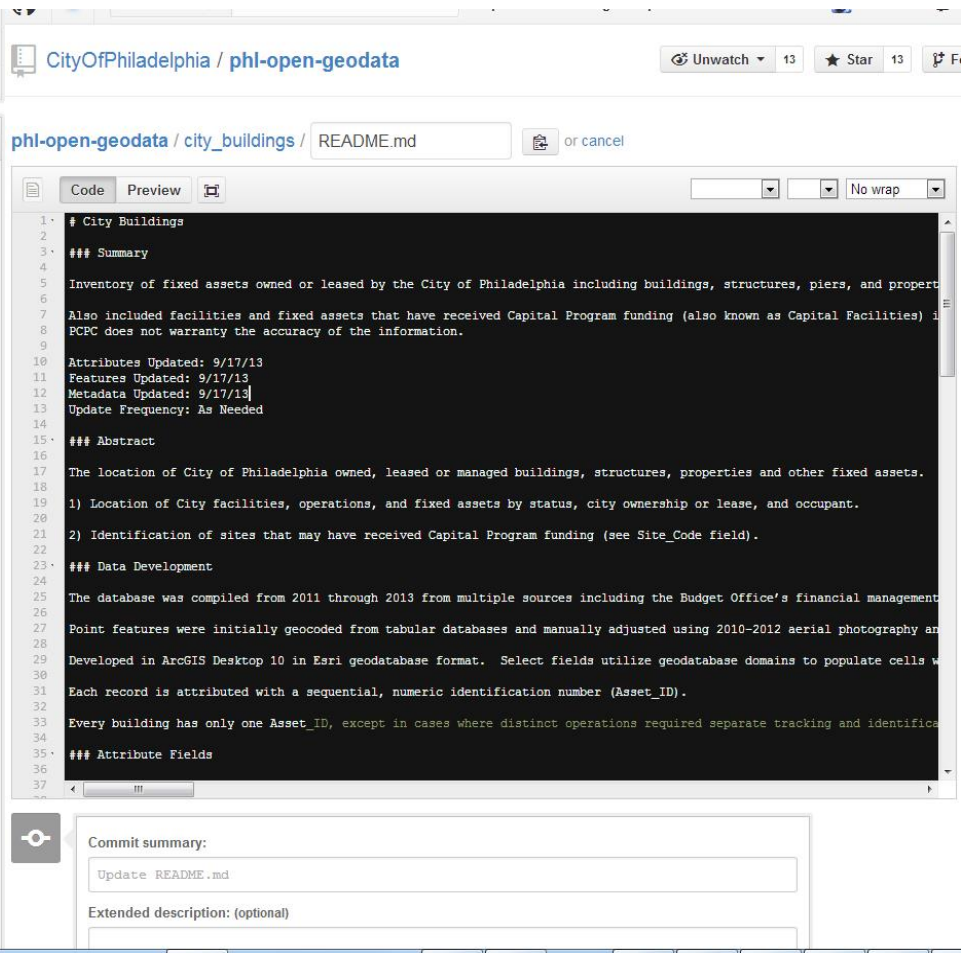
Attributes Updated: 9/17/13
Features Updated: 9/17/13
Metadata Updated: 9/17/13
Update Frequency: As Needed

Abstract

The location of City of Philadelphia owned, leased or managed buildings, structures, properties and other fixed assets. The database includes point feature/records for properties or buildings for which the City has expended Capital Program funds. The primary uses for this feature class:

1) Location of City facilities, operations, and fixed assets by status, city ownership or lease, and occupant.

2) Identification of sites that may have received Capital Program funding (see Site_Code field).



CityOfPhiladelphia / phl-open-geodata

Unwatch 13 Star 13

phl-open-geodata / city_buildings / README.md or cancel

Code Preview No wrap

```
# City Buildings

### Summary

Inventory of fixed assets owned or leased by the City of Philadelphia including buildings, structures, piers, and propert

Also included facilities and fixed assets that have received Capital Program funding (also known as Capital Facilities) i
PCPC does not warranty the accuracy of the information.

Attributes Updated: 9/17/13
Features Updated: 9/17/13
Metadata Updated: 9/17/13
Update Frequency: As Needed

### Abstract

The location of City of Philadelphia owned, leased or managed buildings, structures, properties and other fixed assets.

1) Location of City facilities, operations, and fixed assets by status, city ownership or lease, and occupant.

2) Identification of sites that may have received Capital Program funding (see Site_Code field).

### Data Development

The database was compiled from 2011 through 2013 from multiple sources including the Budget Office's financial management

Point features were initially geocoded from tabular databases and manually adjusted using 2010-2012 aerial photography an

Developed in ArcGIS Desktop 10 in Esri geodatabase format.  Select fields utilize geodatabase domains to populate cells w

Each record is attributed with a sequential, numeric identification number (Asset_ID).

Every building has only one Asset_ID, except in cases where distinct operations required separate tracking and identifica

### Attribute Fields
```

Commit summary:

Update README.md

Extended description: (optional)

Testing

- For GeoJSON – geojson.io
- For KML – Google Earth
- For Markdown – Dillinger.io

Demo!

ArcOpen Scripting

```
import arcpy
arcpy.ImportToolbox('path/to/ArcOpen.pyt')
arcpy.Convert_ArcOpen('path/or/db/connection/to/farmers_markets',
                      'OBJECTID OBJECTID HIDDEN NONE; \
                      NAME NAME VISIBLE NONE; \
                      ADDRESS ADDRESS VISIBLE NONE; \
                      OPERATOR OPERATOR VISIBLE NONE; \
                      ACCEPT_SNA SNA VISIBLE NONE; \
                      ACCEPT_FMN FMN VISIBLE NONE; \
                      DISTRIBUTE DISTRIBUTE VISIBLE NONE; \
                      ONLY_REDEE REDEEM VISIBLE NONE; \
                      EBT_MACHIN EBT VISIBLE NONE; \
                      DAY_TIME DAY_TIME VISIBLE NONE; \
                      ZIP_CODE ZIP_CODE VISIBLE NONE; \
                      SHAPE SHAPE HIDDEN NONE',
                      'c:/github_data',
                      'farmers_markets',
                      convert_4326='true', convert_geojson='true',
                      convert_kmz='true', convert_csv='true',
                      convert_metadata='true', debug='false')
```

Getting Started

1. Create a github.com account (it's free)
2. Get added to the CityOfPhiladelphia organization by emailing me (david.walk@phila.gov)
3. Download the GitHub for Windows application
    a) Optional: Security may have to open port 22 in the firewall for you (OIT can help with this)
4. Download and Install ArcOpen (ArcGIS 10.1 SP1+)
5. Go!

Organizational Structure

- Each department will have one public repo (except for special cases like crime incidents) unless departments want to request a special arrangement
- GIS staff should have their own personal GitHub accounts (free)
- GIS staff will be grouped into GitHub teams by department
- These teams will have pull/push request to their specific repo(s) and no one else will besides the administrators
- OIT will administer the organization

Future

- Detailed instructions created (as a GitHub repo!)
- OIT will **NOT** perform the yearly PASDA update
- All PASDA GIS datasets should be published on GitHub ASAP
- Once a good amount of PASDA datasets on GitHub, OIT will work with PASDA to change links
- OIT will work with Open Data Philly to constantly update links as data is added to GitHub
- Email me (david.walk@phila.gov) with any issues

Resources

- Pro Git book (free)
- Git cheatsheet
- Git Reference
- Try GitHub (free video course)
- 10 More GitHub Tutorials
- Markdown cheat sheet

Thx!

David.walk@phila.gov

https://github.com/CityOfPhiladelphia

https://github.com/CityOfPhiladelphia/arc-open

Questions?