International Cartographic Association
What is open source and how do pay models for access/use/upgrade of open source coexist

Danny Sheehan - Carto
• Solutions Engineer at Carto
  ○ user of FOSS4G tools/code/platforms
  ○ 14+ years other Desktop GIS Software product(s), geoprocessing and spatial analysis.

• Email: danny@carto.com
ABOUT

• Filling in for
  ■ MAMATA AKELLA - SENIOR CARTOGRAPHER at CARTO
About

• Education
  ○ Columbia University
    ■ Certificate in Professional Achievement in Data Sciences
  ○ University at Buffalo
    ■ Masters in Geography/Specialization in GIS
  ○ Geneseo
    ■ Bachelors in Geography
ABOUT

• Prior experience
  ○ **Columbia University** - 6.5 years Senior Research Scientist/GIS Analyst at Department of Epidemiology Mailman School of Public Health - Built Environment & Health Project (beh.columbia.edu),
    ■ Affiliations with Institute of Social, Economic Research & Policy, Fu School of Engineering
    ■ Teaching at Dept. of Epidemiology, Quantitative Methods for Social Sciences, Barnard College
  ○ **AECOM** - 3.5 years lead GIS Analyst for FEIS/FOEIS for US Navy projects, Dahlgren, USWTR, Guam, I-287/Tappan Zee Bridge
ABOUT

• **Journal publications**
  - American Journal of Preventive Medicine
  - American Journal of Public Health
  - European Journal of Gastroenterology & Hepatology
  - Journal of Exposure Science and Environmental Epidemiology
  - Cancer Causes & Control
  - AIDS and Behavior
  - Journal of Maps
  - Etc,...

• **Google School Link**
  [https://scholar.google.com/citations?user=K6iTYsUAAAAJ&hl=en]
ABOUT

• Current Role as Solutions Engineer at Carto:
  ○ On-Prem Install
  ○ Leveraging our APIs and the Carto stack
  ○ Setting up Database Connectors
  ○ Building middleware!!!
  ○ ...(continue)
• Current Role as Solutions Engineer at Carto:
  ○ Bug Reporting and Fixes
  ○ Geoprocessing with PostGIS
  ○ Strategizing analytical methods
  ○ Coordinating between Carto’s Data Science teams and our clients’ Data Science needs
  ○ Getting the Feature Request you ask for into the Carto Platform
What is Open Source?
What is Open Source?
What is Open Source?

Open-source software (OSS) is computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose. [1]
What is Open Source?

Open-source software (OSS) is computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose. [1]

- Wikipedia:

What is Open Data?
What is Open Data?

Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. [1]
What is Open Data?

Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. [1]

- Wikipedia:

1. Desktop vs. Server architecture
1. Desktop vs. Server architecture

2. Installer scripts/UI vs. command line approach/dependencies are up to the installer.
OPEN SOURCE

TYPES OF OPEN SOURCE

1. Desktop vs. Server architecture

2. Installer scripts/UI vs. command line approach/dependencies are up to the installer.

3. Add-on features/plugins only available at paid level vs. everything is open.
Spatial Machine Learning

Leverage features in your own data and location context to train gradient boosted random forests and predict new markets.
LOCATION DATA SERVICES

Exceptional data, technology, and cartography to power your applications.

A set of Location Data Services including vector basemaps, geocoding, routing, and demographic data augmentation services to help extract the full potential of location intelligence.
Data Observatory

**Measurements**
Access demographic, economic, and both numeric and categorical measurements at places of interest.

**Segmentation**
Classifications are built up from leading population data and provide a comprehensive view of where your potential customers, voters, or supporters exist.
RESEARCH

Data Observatory

Big Metadata

https://cartodb.github.io/bigmetadata/
1. Desktop vs. Server architecture

2. Installer scripts/UI vs. command line approach/dependencies are up to the installer.

3. Add-on features/plugins only available at paid level vs. everything is open.

4. License types of Open Source
[https://opensource.org/licenses]
1. Quickly fixable... code often in public repo.
1. Quickly fixable... code often in public repo.

2. More secure... more eyes on the code that can identify security holes/flaws.
OPEN SOURCE ADVANTAGES

1. Quickly fixable... more eyes on the code.

2. More secure... more eyes on the code that can identify security holes/flaws.

3. Active collaboration and community. Which breeds innovation, health debates.
How do pay models for access/use/upgrade of Open Source coexist?
Open Source Models Examples:

From where else than StackOverflow:

How do open source developers make money?

https://stackoverflow.com/questions/8159049/how-do-open-source-developers-make-money
1. The open source project is a side project and have developers have other jobs as their source of income.
1. The open source project is a side project and have developers have other jobs as their source of income.

2. Corporate Sponsorship pays developers and company.
1. The open source project is a side project and have developers have other jobs as their source of income

2. Corporate Sponsorship pays developers and company

3. **OSS project is developed by company and sold services based on that project**
   - (ie, I give away MySQL for free, and do consulting installing and maintaining MySQL for companies)
OPEN SOURCE MODELS

1. The open source project is a side project and have developers have other jobs as their source of income

2. Corporate Sponsorship pays developers and company

3. OSS project is developed by company and sold services based on that project
   - (ie, I give away MySQL for free, and do consulting installing and maintaining MySQL for companies)

4. They develop a base version of the OSS project, and then develop and sell commercial add-ons for it
At Carto seems like we have a combo of all 4

Sponsorship for features from companies, selling of additional or hosted features/storage, we consult, etc. and some contributions are made by our dev. community/our dependencies also have a mix of all four (4).
We’re Open Source built on top of other Open Source
Access/Use, & Upgrade
1. Open Source version, self installed.
1. Open Source version, self installed.

2. Hosted Cloud version (SaaS)
1. **Open Source version, self installed.**

2. **Hosted Cloud version (SaaS)**

3. **On-premise installation licensed and sold by Carto.**
   - Could be either in clients cloud environment, local/virtual on physical server.
1. Storage - clients pay for additional storage.
1. **Storage** - clients pay for additional storage.

2. **Speed/Processing Power**
   - Handle Larger Data Volumes
   - Handle Complex Processes
     - NP Hard, etc.
   - Handle Processes faster
   - Upgrades to core technology/dependencies, i.e. Postgres
1. Storage - clients pay for additional storage.

2. Speed/Processing Power

3. Additional Features
   - Added analysis methods
   - Custom tools
   - Sponsored features
Having these pay models helps support core developers to work on the projects full time (i.e. Red Hat) and improves the project.
At Carto:

Our core development and Product team, Data Science/Research and Solutions team(s) are all individuals that firmly believe in Open Source as a philosophy.

Get a list of all the current import jobs

```python
from carto.file_import import FileImportJobManager

file_import_manager = FileImportJobManager(auth_client)
file_imports = file_import_manager.all()
```

Get all the datasets

```python
from carto.datasets import DatasetManager

dataset_manager = DatasetManager(auth_client)
datasets = dataset_manager.all()
```

Get a specific dataset

```python
from carto.datasets import DatasetManager

# write here the ID of the dataset to retrieve
DATASET_ID = ""

dataset_manager = DatasetManager(auth_client)
dataset = dataset_manager.get(DATASET_ID)
```
Carto Install:

Install:

End, questions?
Contact:
@nygeog
danny@carto.com
BUILDER

LOCATION INTELLIGENCE FINALLY INTUITIVE

A web-based drag and drop analysis tool for business users and analysts to discover and predict key insights from location data.

CARTO Builder unleashes the power of location intelligence with self-service, actionable dashboards you can share across your whole organization.
Intuitive Analysis pipelines

Easily generate powerful analysis pipelines that combine GIS and Data Science workflows.
Useful cross-filtering Widgets

Explore your data on-the-fly
Work in your own Data Science world

CARTO offers CartoFrames a convenient and intuitive Python module that fits into your existing Data Science Workflows and our newly reconfigured Carto Python SDK (new version 1.0.0) that allows you to easily interact with the Carto Engine in your Python environment.
LOCATION DATA SERVICES

Exceptional data, technology, and cartography to power your applications.

A set of Location Data Services including vector basemaps, geocoding, routing, and demographic data augmentation services to help extract the full potential of location intelligence.
Car2Go Examples: Animated Time-Series
**Car2Go Examples:**

<table>
<thead>
<tr>
<th>Date</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-08-09</td>
<td>933</td>
</tr>
<tr>
<td>2016-08-10</td>
<td>908</td>
</tr>
<tr>
<td>2016-08-14</td>
<td>686</td>
</tr>
<tr>
<td>2016-08-12</td>
<td>638</td>
</tr>
<tr>
<td>2016-08-08</td>
<td>609</td>
</tr>
<tr>
<td><strong>OTHER</strong></td>
<td><strong>1.6k</strong></td>
</tr>
</tbody>
</table>

**Name**

<table>
<thead>
<tr>
<th>Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXG4893</td>
<td>73</td>
</tr>
<tr>
<td>AXG4729</td>
<td>57</td>
</tr>
<tr>
<td>AXG5021</td>
<td>48</td>
</tr>
<tr>
<td>AXG5744</td>
<td>46</td>
</tr>
<tr>
<td>AUK2242</td>
<td>46</td>
</tr>
</tbody>
</table>
Car2Go Examples: Car2Go Locations Seattle with Convex Hulls and Connected Lines
Carto | Python

Python is super powerful. Thus, we have our Carto SDK for Carto APIs, new version 1.0.0!

Also, for fans of Python Pandas (R-like operations and data structures) we have newly released CartoFrames.

```python
from carto.file_import import FileImportJobManager

file_import_manager = FileImportJobManager(auth_client)
file_imports = file_import_manager.all()

from carto.datasets import DatasetManager

dataset_manager = DatasetManager(auth_client)
datasets = dataset_manager.all()

from carto.datasets import DatasetManager

# write here the ID of the dataset to retrieve
DATASET_ID = ""

dataset_manager = DatasetManager(auth_client)
dataset = dataset_manager.get(DATASET_ID)
```
Carto-Python is a full, backwards incompatible rewrite of the deprecated cartodb-python SDK. Since the initial rewrite, carto-python has been loaded with a lot of new features, not present in old cartodb-python.

Get a list of all the current import jobs

```python
from carto.file_import import FileImportJobManager
file_import_manager = FileImportJobManager(auth_client)
file_imports = file_import_manager.all()
```

Get all the datasets

```python
from carto.datasets import DatasetManager

dataset_manager = DatasetManager(auth_client)
datasets = dataset_manager.all()
```

Get a specific dataset

```python
from carto.datasets import DatasetManager

# write here the ID of the dataset to retrieve
DATASET_ID = ""

dataset_manager = DatasetManager(auth_client)
dataset = dataset_manager.get(DATASET_ID)
```
Carto | CartoFrames

A Python package for integrating CARTO maps and services into data science workflows.

Examples:
https://github.com/CartoDB/cartoframes/tree/master/examples

- https://github.com/CartoDB/cartoframes/blob/master/examples/Basic%20Usage.ipynb

```python
# Get a CARTO table as a pandas DataFrame
def = cc.read('brooklyn_poverty')
df.head()

<table>
<thead>
<tr>
<th>cartodb_id</th>
<th>the_geom</th>
<th>the_geom</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>01030000020E610000001000000B0000006D3A02B85982...</td>
<td>01030000020E610000001000000B0000006D3A02B85982...</td>
</tr>
<tr>
<td>3</td>
<td>01030000020E61000000100000090000006AA164726A80...</td>
<td>01030000020E61000000100000090000006AA164726A80...</td>
</tr>
<tr>
<td>4</td>
<td>01030000020E610000001000000E0000005922A7AF7F...</td>
<td>01030000020E610000001000000E0000005922A7AF7F...</td>
</tr>
<tr>
<td>6</td>
<td>01030000020E6100000010000008000000F5BD86E0B87E...</td>
<td>01030000020E6100000010000008000000F5BD86E0B87E...</td>
</tr>
<tr>
<td>7</td>
<td>01030000020E610000001000000C000000E6CDE15AED7E...</td>
<td>01030000020E610000001000000C000000E6CDE15AED7E...</td>
</tr>
</tbody>
</table>
```

```python
# from cartoframes import Layer, styling
l = Layer('brooklyn_poverty', color={'column': 'poverty_per_person', 'range': [0, 100]})
cc.map(layers=l, interactive=False)
```

Out[3]:
Spatial Analysis for insights and prediction

It's not just your data that is spatial, the underlying processes that produce it and drive your business are as well.

CARTO offers a suite of analysis methods to understand, quantify and predict those spatial relationships to generate fresh insight and opportunity from your data.
Analysis

Including spatial analysis in your data science workflow can uncover new insights and opportunities

**Optimization**
Analyze your operational infrastructure for inefficiencies and find new optimizations. See how a change will affect your network.

**Inference**
Uncover the hidden relationships in your data to inform decision making, planning and understanding customers behavior.

**Prediction**
Use the spatial and temporal patterns in your data to predict performance in new markets and identify new opportunities for growth.
Intuitive Analysis pipelines

Easily generate powerful analysis pipelines that combine GIS and Data Science workflows.
Work in your own Data Science world

CARTO offers **CartoFrames** a convenient and intuitive Python module that fits into your existing Data Science Workflows

```python
import pandas as pd
def = pd.read_carto(username=USERNAME, api_key=APIKEY, tablename='brooklyn_poverty_notnull_geoms')
def.head()
```

<table>
<thead>
<tr>
<th>cartodb_id</th>
<th>high_school_diploma_2011_2015</th>
<th>median_income_2011_2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>153</td>
<td>0.219213</td>
<td>36414.226756</td>
</tr>
<tr>
<td>946</td>
<td>0.258464</td>
<td>17156.148349</td>
</tr>
<tr>
<td>15</td>
<td>0.264986</td>
<td>45746.135360</td>
</tr>
<tr>
<td>37</td>
<td>0.105692</td>
<td>100039.579057</td>
</tr>
<tr>
<td>1</td>
<td>0.059511</td>
<td>112115.160834</td>
</tr>
</tbody>
</table>

```python
def.carto_map(interactive=True, stylecol='median_income_2011_2015')
def.carto_map(autoscale=True)
```
Inference

Your data contains underlying patterns and relationships that vary from place to place.

Use CARTO’s inference tools to uncover them and use them to make better decisions.
ANALYSIS

Outliers and Clusters

Find clusters of high or low values and also outliers by measuring Spatial Autocorrelation.

Clusters to tweets mentioning earthquakes during an unusual quake in Kansas in 2016
Geographically Weighted Regression

Uncover the Relationships In your data and how they vary spatially.

Relationships between educational attainment, poverty and diversity in rural and urban environments.
Most infrastructure is spatially distributed and incurs costs. Use CARTO to profile and configure your operation to minimize costs and identify key high nodes within your operation.
ANALYSIS

Linear programming

Use our linear programming module in conjunction with Data Locations Services to optimize and fine tune your decision making.
Region Definition

Use our Max-P and spatial Agglomerate Analysis Methods to define balanced sales regions, distinct markets and more.
Use your existing data to predict how your business will perform in new markets and how existing markets will evolve.
SPATIAL MARKOV CHAINS

Use the past performance of your locations and their neighbors to predict trends in the market.