May 9 2017 – Eve Ahearn

ThoughtWorks PASSCODE:
ID: TWGUEST
PASSWORD: "begin hark sauce editor sign"

NOTES: http://bit.ly/phn9may
Sign up for slack: proghacknight.org

Progressive HackNight is an intersection of technology and activism. A physical space for progressive activists, organizers, ux/ui designers, and coders to come together, discus the issues, and work on projects together that will drive the movement forward.

The speaker for the next hacknight is Eve Ahearn, Product Manager at Enigma.com. Eve is scouting out cool data. She completed a MSc at the Oxford Internet Institute, thinking about data policy under the supervision of Viktor Mayer-Schönberger. In a past life, Eve spent a year teaching AP U.S. History at Weston Prep in Panama City, Panama.

Food will be provided!

The schedule is as follows:

6:30p - 6:50p : registration, networking (mingle!)

6:50p - 7:10p : introductions, announcements, discussions

Announcements

- Make The Road – Backers of Hate
- 10 Ways Coders can help the fight for Universal Health Care

7:10p - 7:30p : Speaker

7:30 - 7:40p : Breakout introduction

7:40p - 9:30p : Breakout

Breakout Groups

If you have a breakout group idea, start one by adding it to the ProgHackNight github. There are two types of breakout groups: discussion group and working group.

DISCUSSION GROUP

Groups centered on learning new things, discussing issues, building networks, and strategizing. Start one through GitHub or Google Form (we'll put it in GitHub)

WORKING GROUP

Working groups are breakouts focused on either onboarding, training, or just having a team setup on collectively coding for your awesome project! Start on via GitHub or Google Form.

In Solidarity,

The ProgHackNight team

NOTE TAKING: Eve Ahearn – Enigma.com

Eve – Product manager for Enigma Public, repo for public datasets = have 1000s of datasets, primarily US, but hoping to change that. Launching new version next month; main change is to make data more accessible, so getting rid of login to access data (don't need to create an account now).

1. What is public data?
    a. Not all of the data we use is open
    b. Released by governments, data released by academics
    c. It's not open data, we have to submit FOI data
    d. Buy a lot of data in CD and hard-drive.
    e. For the datasets we buy, we're the only one that is free.
    f. NYC Restaurants, OSHA

2. What's Enigma
   a. Data management system, ETL ecosystem
   b. A lot of our clients are data scientists

3. What's Enigma Public
4. Demo
5. What's covered?
6. Labs - Healthcare and sanctions
7. What's next?

Sanctions project: started to track Trump's actions on sanctions and whether they were in line or out of line with previous administrations.

Public data?

Is data that's in the public domain. Eve did a lot of stuff with open government data… know about open data, open government data, but what is "public data". Why "public data"? Because not all the data used is open… public is in the public domain, e.g. has no restrictions on reuse… could be released by govts, academics, nonprofits, some companies. Company does a lit with open data - love for more data to be open - gets a lot of data from FOAIs, buys a lot of data from govt agencies, in every format imaginable… for datasets bought or FOIAs by enigma, are the only place can get this data for free online.

Examples: OHSA, NYC restaurant data.   Example: Empire State bldg has own zipcode… can explore site by address, looked at different datasets by zipcode for this, e.g. for just one building. For a building have e.g. corporate registrations: companies their officers, FCC donation records (NB usually people put down their home address), elevator inspection records, building permits etc.

Enigma

Does a lot with public data: data operations and infrastructure, not data analytics company. Build ETL etc; lot of tools to get data into a state so data scientists can do their work.

Enigma public = free community application. Client = intersection of client internal data with public data. (As usual, they're hiring; currently 75 people, midtown NYC).

Assembly = data mgmt; concourse = etl system; help companies use data (companies like it because data is accessible); all data is released for free under creative commons licenses on enigma public.

Demo

Old version of site is https://app.enigma.io - will have to create an account. Will look very different in about 4 weeks' time.

Can search site: both the contents of the data, but also table descriptions and metadata. One thing difficult with govt data is finding it, even if has a web portal; interesting stuff is often held on website as reports… even in NYC, have data at state, county and city level; hard to think of all the possible datasets you could use that relate to a place. Example: search for Google.

Can also browse through data, e.g. by topic. Example: FCC campaign contributions 2016. Can use API or export the data, but also have quick summary statistics (max, min, avg, sd etc) for each dataset before diving in deeper.

Enigma public:

Looking at http://www.enigma.com/public - contains guides for how to use specific datasets.

Why people use enigma public instead of source? Not all data is available or easy to get online (e.g. those FOIAs; have OPM dataset en route to them now - federal govt employees, positions, salaries etc). Can be difficult to find datasets on a specific subject. Reliability: updating data so you don't have to return to source (and get notified when sources change).

Cool projects

The sanctions tracker (Eve worked on this). Tracks changes to sanctions programs, e.g. changed today: people removed from list, e.g. narcotics trafficking; biggest change was Trump adding 300 Syrian scientists to the list 3 weeks ago (PR? People don't necessarily have access to US banking)... mapped this, compared across different presidents (Bush, Obama etc), administrations (e.g. Myanmar) and programs (terrorism, narcotics etc).

Healthcare data project: intersection between patent applications and cost of drugs in medicare. Patent cliff = when drugs go off patent; see companies re-patenting a very similar version of a drug right before the patent cliff (e.g. it was a pill, they make a spray instead). USes datasets from 12 different agencies, on prescriptions.

Getting started with Enigma

1. Start exploring enigma public
2. Create your enigma public account
3. Sign up for the newsletter (coming soon!)

Eve will send update to Slack group after relaunch; newsletter will start then too (people using datasets on projects, new datasets etc in that).

Q&A

"If I was an academic researcher, I'd probably not use Enigma". We do have academics who use the site (encouraged to use free version) rather than customers. Work with a lot of journalists too. Even in academia, people have different levels of skills, e.g. social science researchers wanting to use data but aren't technical. Even if technical, building a pipeline takes time: people use data from enigma first to see if data work working with before building pipelines.

"Enigma public: have you seen people using data in interesting ways for e.g. social justice". Y, but want to see a lot more of that.

"We're a social justice org using a lot of data - govt datasets, FOIAing payroll records at state/local level, enjoying cleaning that data. How much is Enigma public a space to collaborate on FOIAs?". In past, have relationship with Muckrock platform (makes it easy to submit FOIAs, can set to e.g. delay making it public for journalists for a while so they can get story out), have collaborated with them in past. Want after relaunch to be more proactive about partnering with people, want to hear about requests for data; let's talk about things you want to FOIA for.

"Concerned that data find useful, are crossing variables with each other; do you provide a data analysis tool too?" Company not about data analytics tools - more about data. People use a variety of tools for this.