

HEALTH EQUITY HACKATHON

Introduction	1
Data Set By Problem Area	3
Addiction	3
Obesity	4
Access to Care	4
Women’s Health	4
Social Determinants	5
Socrata End Points	5
Unique Data Sets	6
Data Set Logistics	7
Participant Qualifications	8
Outcome Goals	9
Potential Data Projects	10
Proposed Problem/Project Pairs	11
Hacking Where?	11
Making Access to Knowledge Accessible	11
Kratom Evidence?	11
Everyone Can Hack.	11
Wikipedia Medical Reference Editing Tool	11
Modeling healthcare related Wikipedia projects	11
Nobel Chores	11
Hospital Ceilings	11
Wikipedia Hacking projects.	11
Hashtag	12

Introduction

Dates

October 19-21 | Austin, TX

The Health Equity Hackathon is about: techies and community members coming together to overcome systemic barriers to building healthy communities for marginalized populations.

Most people assume that Hacking only means breaking into computers. But the word “Hacking” is a technorati [shibboleth](#). Those who are deeply informed in technology know that the original meaning of the word originally meant something like “The creatively overcoming or circumventing limitations that probably should not exist.”

This Hackathon is all about collaborative applying the original hacking ethos to the most difficult problems in healthcare, especially those caused by poverty and discrimination and result in significant disparities in health access, treatment and outcomes. Our focus for this inaugural event will be broad yet focused on Addiction (Opioid Crisis, Alcohol, Tobacco), Obesity, Women’s Health and Health Care Access.

More importantly, the Hackathon will approach these issues from a health equity lense, which includes:

- A commitment to building *healthy* lives and communities, not just *medical care*
- Participation from the communities most directly impacted
- A focus on the specific (demographic and geographic) populations with who have been adversely affected
- Attention to the systemic barriers and root causes, rather than individual behavior change or “band-aid” solutions
- Removing social and economic obstacles, such as poverty and discrimination
- Reducing and eliminating disparities in health and their determinants:
 - Economic stability and employment
 - Education
 - Food (and water) access
 - Health Care Access
 - Housing and the physical environment
 - Social bias and supports

Basic principles of how this hackathon will be different/better than standard hackathons.

Overall, the right way to think of this Hackathon is “barn building not a beauty contest.” These are deep-rooted, complex challenges that require thoughtful, sustained, collaborative commitment and oftentimes simple solutions. A great hack is very frequently simple, boring and difficult. In many cases, healthcare hacks are just chores. Sometimes these changes do not solve problems directly, but instead make other innovations easier. Sometimes they just help a little. This hackathon will celebrate doing a good job on small-but-important tasks, as equal to the profound insights that are typically celebrated as innovations. Guiding principles include:

- ***Focus on solving a health equity problem***, rather than building a startup or an “app.” Specific focus areas:
 - Addiction (Opioid Crisis, Alcohol, Tobacco)
 - Obesity
 - Women’s Health
 - Health Care Access
- ***Not like a race or a beauty pageant, more like a barn building.***
 - No prizes for winning ⇒ resources to solve problems.
 - No judges ⇒ mentors.
 - Less competition ⇒ more collaboration
 - Community-driven ⇒ not “top-down” or “outside-in” solutions
 - Medtech experts ⇒ cross sector skills and experiences, including community members
- ***Invite specific collaborators and mentors and use them to draw the other talent***, including: “experts” from across multiple sectors (not just tech and medicine) and community members who are closest to the issues. Qualifications are based on personal and professional skills, knowledge and experience.
- ***Focus on preparing “goldilocks problems.”***
 - Hard enough to interesting. Real problems.
 - Simple enough that demonstrable progress can be made with real teams
 - Problems that focus on match-making... I.e. “if we had a developer and a designer. We could (solution goes here)”
 - Build on the strengths and resources already active in communities (people, stories, economic, physical, associations, organizations) to ensure sustainability
- ***What not to do***
<https://www.wired.com/story/sociologists-examine-hackathons-and-see-exploitation/>

Data Set By Problem Area

Notes on the wikipedia datasets

All of the wikipedia datasets have three pillars:

1. <i>Wikipedia article pageview data</i>	https://dumps.wikimedia.org/other/pageviews/
2. <i>Wikipedia click stream data</i>	https://dumps.wikimedia.org/other/clickstream/2018-08/
3. <i>Wikidata on specific pages.</i>	(these are the infoboxes on the pages.)

Good places to start for wikipedia data are the:

<i>Statistics page:</i>	https://en.wikipedia.org/wiki/Wikipedia:Statistics#Datasets
<i>WikiData page</i>	https://www.wikidata.org/wiki/Wikidata:Main_Page

Datasets by challenge

Addiction

MANY DIFFERENT DRUGS	
<i>Erowid-</i>	https://erowid.org/ thousands of illegal drug experiences chronicled here.
<i>Reddit -</i>	There is a subreddit for many different opioids including Kratom
<i>Poison Control dataset (stuck in pdf)</i>	https://www.aapcc.org/annual-reports
OPIOID	
<i>Kratom dataset</i>	released by CareSet at hackathon. No link
<i>Part D Utilization data</i>	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.html
<i>Wikidata has Opioid Receptor mode of action</i>	https://en.wikipedia.org/wiki/Opioid_receptor
ALCOHOL	
<i>ATF Statistics and Data</i>	https://www.atf.gov/resource-center/data-statistics
<i>Liquor License data (sporadically available per city/state)</i>	https://catalog.data.gov/dataset/liquor-licenses-5a0dc https://data.ny.gov/Economic-Development/Liquor-Authority-Quarterly-List-of-Active-Licenses/hrvs-fxs2
METH	
<i>Data behind these articles</i>	https://www.ncbi.nlm.nih.gov/pubmed/18472624 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4440680
<i>Hospital Admission data (start here)</i>	https://www.samhsa.gov/sites/default/files/teds-short-report043-urban-rural-admissions-2012.pdf
<i>DAWN Drug Abuse Warning Network</i>	https://www.datafiles.samhsa.gov/study-series/drug-abuse-warning-network-dawn-nid13516
<i>TEDS Data</i>	https://www.dasis.samhsa.gov/webt/newmapv1.htm
FOOD	

<i>Nutrition Survey - NHANES 2013-2014</i>	https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013
<i>FoodGraph Data -</i>	https://foodgraph.docgraph.com/
<i>Wikipedia Nutrition and related articles</i>	https://en.wikipedia.org/wiki/Human_nutrition
<i>Article about food security apparently good work being done from UCSF</i>	https://sirenetwork.ucsf.edu/tools-resources/resources/implementation-food-insecurity-screening-and-referral-program-student-run
<i>Connections between SNAP benefits and food insecurity</i>	

Obesity

<i>Multiple life/death datasets under "Life Stages and Populations"</i>	https://www.cdc.gov/nchs/fastats/life-stages-and-populations.htm
<i>Specifically causes of death</i>	https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm
DIABETES	
<i>83036 CPT Code (H1Ac test order) in Part C data and Medicare Fee for Service data.</i>	(Part C util file will be released by CareSet at the hackathon) . Does not show test results but does show ordering rates...
<i>CDC Diabetes data</i>	https://www.cdc.gov/diabetes/data/
CARDIOVASCULAR DISEASE	
<i>CDC heart disease data</i>	https://www.cdc.gov/heartdisease/maps_data.htm
<i>Stroke</i>	

Access to Care

<i>Classic dataset is the Dartmouth Atlas, which is GEO coded and derived from CMS data</i>	https://www.dartmouthatlas.org/tools/downloads.aspx
<i>CDC access to care</i>	https://www.cdc.gov/nchs/fastats/access-to-health-care.htm
<i>FQHC data</i>	https://bphc.hrsa.gov/datareporting/
<i>National Association of Free Clinics data</i>	https://www.nafcclinics.org/advocacy/tools
<i>FQHC lookalike data</i>	https://bphc.hrsa.gov/uds/lookalikes.aspx

Women's Health

<i>NIH Office of Women's Research inclusion</i>	https://orwh.od.nih.gov/toolkit/application-reporting-process/reporting-inclusion-enrollment-data
<i>Women of Color health data book (in a pdf)</i>	https://orwh.od.nih.gov/sites/orwh/files/docs/WoC-Databook-FINAL.pdf
<i>CDC women's health data</i>	https://www.cdc.gov/nchs/fastats/womens-health.htm
<i>Planned Parenthood annual report</i>	https://www.plannedparenthood.org/uploads/filer_public/71/53/7153464c-8f5d-4a26-bead-2a0dfe2b32ec/20171229_ar16-17_p01_lowres.pdf
<i>Catholic pregnancy centers?</i>	https://www.catholiccharitiesusa.org/financials/
<i>Perhaps Form 990 data</i>	https://registry.opendata.aws/irs990/
<i>HPV Vaccine</i>	https://www.cdc.gov/vaccinesafety/vaccines/hpv/hpv-safety-faqs.html https://www.who.int/bulletin/volumes/85/9/06-038414/en/
<i>Traditional data sources and research</i>	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5473416/

Social Determinants

ICD Z CODES	
<i>ICD 10 Z from Medicare</i>	ICD Z codes are social determinants of health, they are being adopted slowly. Should CareSet make this?
<i>Social Determinants of Health in EHRs</i>	https://www.healthit.gov/sites/default/files/facas/HITPC_Capturing_social_behavioral_domains_and_measures_EHRs_HIT_Policy_2014-12-09.pdf https://www.3mhisinsideangle.com/blog-post/icd-10-adds-more-detail-on-the-social-determinants-of-health/
GEO CODING OF HEALTH RESOURCES	
<i>Community health profile in NYC by neighborhood</i>	https://www1.nyc.gov/site/doh/data/data-publications/profiles.page#bk
<i>Dartmouth Atlas under Access to Care.</i>	
<i>Prevention Status Reports CDC</i>	https://www.cdc.gov/psr/
<i>Zillow</i>	Multiple regional datasets https://www.zillow.com/research/data/Neighborhood-boundary-data https://www.zillow.com/howto/api/neighborhood-boundaries.htm
CRIME DATASETS	
<i>Crime</i>	https://catalog.data.gov/dataset?tags=crime These are typically released on a per-city basis and many of them are cataloged here.
<i>Houston (not included in the catalog)</i>	http://data.houstontx.gov/dataset/houston-police-department-crime-statistics
<i>CDC</i>	https://www.cdc.gov/dhdsp/maps/quick-maps/index.htm

Socrata End Points

Socrata has multiple massive government data sources that they host, at multiple levels of government. These include:

<https://data.cms.gov/>

<https://healthdata.gov/>

<https://www.data.gov>

<https://opendata.usac.org/>

<https://data.cdc.gov/>

<https://data.medicare.gov/>

<https://openpaymentsdata.cms.gov/search>

<https://health.data.ny.gov/>

<https://data.oregon.gov/>

<https://dev.socrata.com/foundry/health.data.ny.gov/353x-fa7w>

<https://health.data.ny.gov/en/browse?q=opioid>

<https://github.com/>

Unique Data Sets

The data list by problem is a better list of third-party datasets and how they might relate to challenges. This list includes the datasets that are going to be specifically made available at this hackathon, several of them for the first time ever. Further, for these datasets, (which will remain in flux until the day of the conference) we will make special efforts to ensure are in an “already downloaded” state.

All data will be made available either via Socrata/Google Cloud. We will also try to have local USB drives.... And likely resilio sync links <https://www.resilio.com/individuals/>

Provider	Data Set	Data Description
CareSet	Regulations.gov Kratom Comments dataset	Background on regulation , a huge number of comments that the public has made on the use of these opioids in response to regulations...
CareSet	NDC Code Pile	There is a lot of chaos in the NDC code space. There is a hacking project that started at datapalooza to fix this. It's boring but important.
CareSet	Medicare Advantage (part C) Utilization	National and state level HCPCS utilization
Wikipedia	Links between healthcare articles	How Healthcare articles link together and to SNOMED/ICD/etc ontologies
Wikipedia	Travel between wikipedia articles	How often people click between wikipedia articles
Wikipedia	Internet in a box	The subset of wiki project medicine that is included in the internet in a box devices.

Data Set Logistics

Everyone downloading multi-gigabyte csv's at the same time on the same wifi, can slow progress to a halt on the first day of the event. We will take three steps to avoid this:

1. ***All of the datasets loaded on multiple (10 or so) usb sticks.*** This would be enough by itself, but people are justifiably reluctant to use USB sticks given recent vulnerabilities.
2. ***Datasets loaded on a laptop, and shared using file sharing systems.*** Bittorrent works, Resilio Sync and [SyncThing](#) should all be configured (we may want to reduce to just one of these)
3. ***All the database loaded into Google BigTable and/or AWS Aurora.*** So that people can quickly and easily get to the data in the cloud without downloading at all.

If for some reason we cannot get these methods ready for a dataset... then we may make the hard decision not to use that dataset.

Participant Qualifications

The most important qualification for participation is enthusiasm! Any person who cares about these types of issues who is capable of high-school level reading, using excel or using a web-browser can contribute. We have many valuable tasks that do not require extensive expertise.

However, if you have any of the following qualifications they are especially helpful. We have many complex tasks that require collaborations between the experts of the following types. Do not worry that you are “only” one kind of expert, we will try to match you with people who can collaborate to meet gaps.

- **Problem Owners** - clinicians patients, public health experts and anyone else who might be able to tell if a give solution might be effective.
- **Clinicians** - people who can evaluate the clinical goodness of proposed solutions and who are preferably near the problem
- **Programmers/Data scientists** - people who know how to work with data, ETL, dashboards, scrapping, natural language processing, hardware-hacking, sysadmins - all geeks welcome.
- **Designers** - people with experience in design thinking
- **Students** - Focus on college juniors and seniors (not turning freshman/sophomore away), and graduate students with an interest in these topics.

Tech Meetups & partners we'd like to invite:

- [Austin AI Group](#)
- [Austin Urban Technology Movement](#)
- [Black Tech Women](#)
- [Dallas Data Science Applications](#)
- [Freetailhackers](#)
- [Hispanic Hackers](#)
- [Houston Data Science](#)
- [LGBTQ in Tech](#)
- [Net Squared Houston](#)
- [Net Squared San Antonio](#)
- [Non profit Tech Club Austin](#)
- [Texas Convergent](#)
- [Urban Co-Lab](#)
- [Women in Computer Science - HornsLink - Campus Labs](#)

Health/ Health Equity/ Community partners we'd like to invite:

- [Austin Black Physicians Association](#)
- [Hispanic Health Professions Organization](#)
- [Texas Health Institute](#)
- [Center for Place-Based Initiatives | Dell Medical School](#)
- [Central Health Home - Central Health](#)
- [Austin Public Health](#)
- [Harris County Public Health](#)

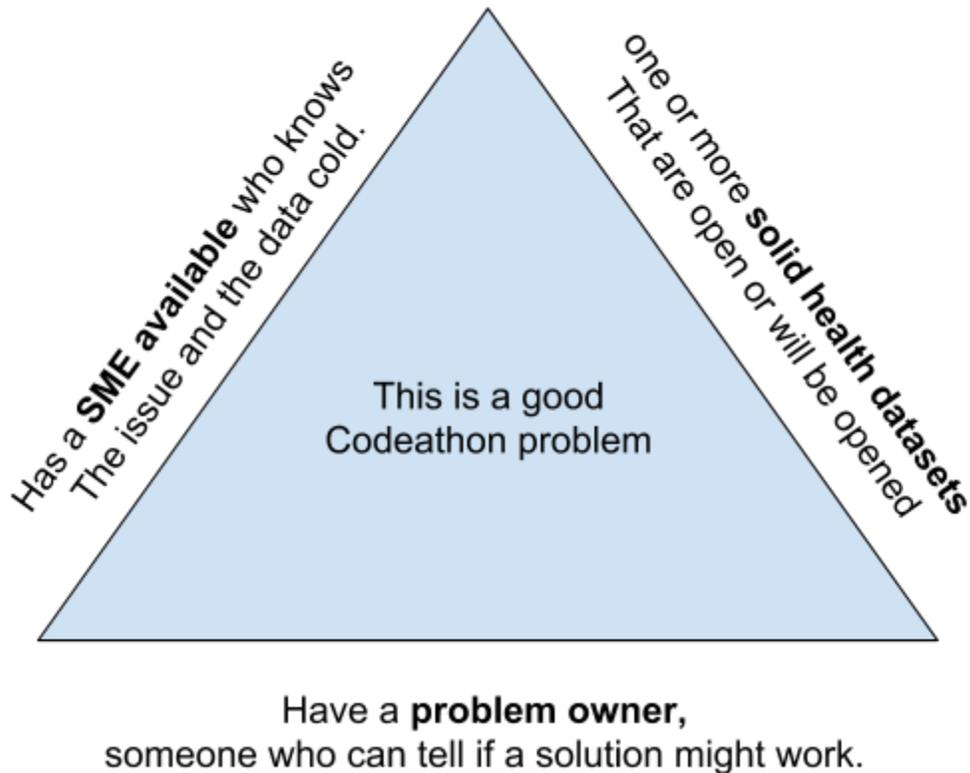
Outcome Goals

This is intended to be a barn-building style hackathon. An ideal solution would balance:

<i>Product</i>	<ul style="list-style-type: none">● address health equity pain points● Address health-related, not just medical solutions● Address social determinants or “root causes”● Focuses on a specific demographic or geographic community● has “real-world’ application● better than toys (i.e. things that are fun but do not matter)
<i>Process</i>	<ul style="list-style-type: none">● still doable over a weekend. By “doable” we mean that a clear progress line can be reached. This does not necessarily mean 100% done, it can mean a prototype, proof-of-concept, or really any reasonable but measurable “short sprint outcome”● prefer to release open resource (open source software github releases wikipedia resources)● was developed with community input (those closest the issue)
<i>Relationship</i>	<ul style="list-style-type: none">● is or can connect to an appropriate community, agency or industry partner that can keep the momentum going

Potential Data Projects

Good problems for the hackathon look like this:



[Data Project Triangle List](#)

Proposed Problem/Project Pairs

Hacking Where?

Can we create resources for entrepreneurs and public health officials to better locate resources in order to impact location based problems?

Making Access to Knowledge Accessible

Can we improve Wikipedia's ability to deliver clinical content? In other languages? To other cultures?

Kratom Evidence?

Can we better understand how grey market drugs (specifically Kratom) are impacting the Opioid Crisis? What about alternative data sources, like Erowid <https://erowid.org/> and

Everyone Can Hack.

Teaching how to contribute to Wikipedia projects, specifically adding reference to medical articles
Teaching how to contribute to Mozilla projects, specifically recording reading snippets

Wikipedia Medical Reference Editing Tool

Help Wikipedia editors by developing a chrome or firefox extension that allows for the reverse lookup of PubMed references, and shows what other articles link to that reference.

Modeling healthcare related Wikipedia projects

Nobel Chores

There are some tasks that are simply complex ETL tasks.. Things that have great value to healthcare informations, but that are kinda boring data munging, data analysis or even data entry tasks that would nonetheless really help mankind. These include (in no particular order)

- Mining the Form 990 data to comprehensively map Form 990 to Hospital NPI information <https://registry.opendata.aws/irs990/>
- Helping to model NDC codes to brand names correctly. "NDC pile project".
- Helping to check that references on Wikipedia medicine articles are actually references that support the sentence in question.

Hospital Ceilings

There is a twitter thread that indicates this is a poorly designed part of the hospital experience.

<https://twitter.com/GraceCordovano/status/1045332884515672066>

Wikipedia Hacking projects.

There are multiple wikipedia hacking projects start looking here:

<https://github.com/Daniel-Mietchen/ideas/issues?utf8=%E2%9C%93&q=is%3Aissue+is%3Aopen+label%3Ahackathon>

More direct link: <https://github.com/Daniel-Mietchen/ideas/issues/966>

Hashtag

#HackingHealthcare

Who should be tagged to get good retweets?

@fredtrotter

@careset

@thoughtleaderRN

@NormaPadron_

@thoughtleaderRN

@McCarthyChris

@mackert

@MonaSiddiquiMD

@nickisnpdx

@Stacey_Chang

@Susan_Hutfless

@ushahmd