

University of Chicago Civic Data and Technology Clinic PIT-UN Year 2 Challenge (2020)

Goals and Objectives

The launch of the Civic Data and Technology Clinic at the University of Chicago in 2020 with the support of PIT-UN funding has been an exciting opportunity to strengthen the university's commitment to the developing field of public interest technology. This seed funding has been critical to the success of the Clinic course, which has run every quarter since Autumn 2020 and has had an enormous impact on students and public interest organizations alike.

The Civic Data and Technology Clinic is now an established part of the public interest technology ecosystem at the university and is actively partnering with public interest organizations to leverage data science research and technology to address pressing social challenges. The Clinic is providing students with exposure to real-world projects and problems that transcend the conventional classroom experience. Driven by the need for innovative training in data science education, the Clinic is an experiential project-based course where students work as data scientists with real-world clients under the supervision of faculty experts.

The Civic Data and Technology Clinic set out with three primary objectives, which have each been met and are detailed below:

- 1. Partner with public interest and government organizations to help them gain insights from their data, innovate, and experiment with proofs of concept, as well as develop software tools to solve societal problems.**

To date, the Clinic has partnered with 13 public interest and government organizations including:

1. Blue Ocean Gear
2. Center for Biological Diversity
3. Center for Research Informatics - UChicago
4. Chicago Human+AI lab
5. Fermi National Lab
6. FracTracker Alliance
7. Hohonu
8. Inclusive Development International
9. 11th Hour Project mBio NLP
10. City of Chicago Office of Mayor
11. Remora Fishing
12. Safety Net Tech
13. Schmidt Ocean Institute

These organizations have a wide range of social impact missions from environmental conservation, to digital equity, to building healthy communities.

2. Empower those organizations to use evidence-based and data-driven approaches to expand their impact.

Partner organizations have joined the Clinic with a wide-range of questions and problems and have been eager and enthusiastic to work with students, faculty, and staff towards making progress on their goals and serving their communities. In the first two quarters of the Civic Data & Technology Clinic, students worked with organizations such as Hohonu, which provides coastal monitoring that assists communities harmed by frequent flooding, and Inclusive Development International, which partners with local communities on economic justice protections. Projects drew upon data science, data visualization, machine learning, and other approaches to build an application to track palm oil deforestation, a system for informing the public about oil and gas industry effects on local air quality, and a real-time dashboard for predicting water level changes across the country.

With FracTacker Alliance, students turned a one-way data path where Colorado residents submitted complaints about the oil and gas industry into a multi-pronged system that communicates rich information to regulators and back to the public. For each complaint submitted, the app built by the project team creates automated reports, attaching numbers and visualizations about the air quality in the user's area and nearby facilities which may contribute to pollution. The reports are then e-mailed to the Colorado Oil and Gas Conservation Commission and the original complainant.

These are just a few examples of successful projects completed through the Civic Data and Technology Clinic. We look forward to increasing the impact and scaling this work in the future to new organizations. A brief synopsis of remaining organizations and clinic projects from the past year is included below.

Blue Ocean Gear - Remote Ocean Buoy Sensor Monitoring

Fishing gear accounts for roughly 10% of all man-made debris in the ocean. Between 500,000 to 1 million tons of fishing gear are discarded or lost in the ocean every year. Blue Ocean Gear deploys remote sensing devices to monitor fishing equipment used by the commercial fishing community. These devices help fishing boats track and make better use of their gear, while also helping to reduce lost fishing waste. In addition to tracking the location of fishing gear, these devices provide valuable measurements such as the depth of the line, the temperature of the water, and the velocity/acceleration of the buoy. Using these measurements the clinic team developed models to identify anomalies in fishing gear readings to help the fishing community track their gear and prevent fishing waste.

Center for Biological Diversity - Ocean Acidification Automated Monitoring

The oceans absorb carbon dioxide pollution and as a result they are becoming more acidic. Ocean acidification makes it difficult for marine animals to build the shells and

skeletons they need for survival. The Clean Water Act provides a tool for states and the EPA to identify and address ocean acidification. Every two years, states identify water bodies that are impaired, and EPA has directed states to consider ocean acidification impairments during their biennial reviews. During these reviews states work with data that has been submitted to their regulatory agencies. In this clinic project, students built a data pipeline for collecting, cleaning, and submitting data from research projects and government agencies so states have the most accurate information for their review of ocean acidification.

Center for Research Informatics, UChicago - National Trauma Data Bank® (NTDB®)
The National Trauma Data Bank (NTDB®) is the largest aggregation of U.S. trauma registry data ever assembled. This includes hundreds of thousands of patient encounters in the context of trauma. We plan to extract data relevant to severe traumatic brain injury and explore variables relevant to outcomes following severe traumatic brain injury. The goal was to ultimately train and validate a predictive model that can assist in selecting candidates for surgical management of severe traumatic brain injury vs medical management.

Chicago Human+AI Lab - Cross-Community Toxic Comments Detection

Toxic comments are disruptive and harmful to online communities. Despite the importance of detecting toxic content, it is very hard to detect new types of content in practice because of the varying definitions and standards across platforms and communities. The goal of this project was to build a dataset for cross-community toxic comment detection by collecting datasets of removed comments from a diverse group of Reddit communities as well as rules information of these communities. Students also started building NLP models for this task.

Fermi National Lab - Using MicroBooNE Radioactive Source Data to Develop Deep Learning Based Triggering for DUNE

The future Deep Underground Neutrino Experiment (DUNE) aims to detect neutrinos produced from the sun and from supernova explosions from the Milky Way or nearby galaxies. The MicroBooNE LArTPC at Fermilab plans to conduct a series of special runs with the injection of radioactive radon into the detector, producing a chain of low energy signatures that can be used to develop and demonstrate real-time identification algorithms will work in future large scale neutrino detectors such as DUNE. The student team identified, developed, tuned and applied algorithms based on ML, deep learning, CNNs and computer vision to MicroBooNE LArTPC data in order to identify low energy signatures in real time.

City of the Mayor (Chicago) - Broadband Disparities Mapping

The goals of this project were to summarize who has broadband at home and who doesn't, determine why access is limited (is it an affordability or infrastructure problem?), and identify what variables influence an individual's likelihood of having broadband at home. Students looked for compiling and mapping trends using various sources to paint

a picture of broadband access. The findings ultimately informed the City's broadband equity investments and initiatives over the next year.

Inclusive Development International (IDI) - Palm Watch

IDI partners with grassroots organizations and local communities around the world to defend their land, natural resources and human rights against threats from harmful investment projects. The goal of the project was to develop a "palm oil tracker"—a dashboard linking deforestation to mills and associated consumer brands. Students used the latest lists of palm oil mills published by the brands, extracted the data from those documents, and then performed record linkage with known mills on the Universal Mill List (UML) to yield a more accurate picture of the palm oil supply chain. They then updated the front-end to allow users to select and search for regions, view region statistics, and see the region and its mills displayed on the map.

mBio - NLP Analysis of African Biotechnology Media Coverage

The 11th Hour mBio Project is a collaboration between the University of San Francisco, the University of Cambridge, and UChicago. mBio works at the nexus of energy, food & agriculture, and human rights. Since at least 2000, a global network of private industries, development agencies and philanthropic donors have promoted biotechnologies on the African continent; now nearly a dozen countries have genetically modified organisms (GMOs) in some stage of research or commercialization. African civil society organizations (CSOs) have raised important critiques, noting that biotechnologies pose threats to African sovereignty and the environment. They have called for greater transparency from the institutions ushering in these new technologies, as well as monitoring of the global flow of finance sponsoring this work. mBio works to build public datasets and analytical tools which empower civic engagement and deepen social impact on a global scale. The clinic project focused on analysis of two million African media articles and their coverage of biotechnologies using natural language processing.³ The goals were to update quotation extraction and analysis pipeline using coreference resolution and other NLP techniques as well as expand, retrain, and fine-tune existing NLP models for sentiment analysis, entity recognition, quotation extraction, and topic modeling. The students then analyzed African biotechnology networks using citation and funding connections.

Remora - Fishing Traceability

Remora develops and deploys computer vision solutions for fish processing facilities in Costa Rica. They collect video from the local fishing community as they process their catch. The platform they have developed records both the species and the weight of the fish.

During this clinic project, the students worked to automate both of these data points using training data generated by Remora's platform. This included volumetric estimation of fish being processed and species classification.

Safety Net Tech - Visualization of 3D Ocean Parameters

SNTech's latest products aim to convert fishing vessels into sophisticated ocean data

collection platforms to create mass-scale subsea ocean measurement. The ocean surface can easily be measured from space, but measuring underwater is expensive and difficult, requiring in situ devices. The subsea environment is poorly understood, negatively affecting our climate models, weather prediction and fishing activities. The project aimed at taking the ocean data generated from fishing vessels and turning it into actionable insights.

Schmidt Ocean Institute (SOI) - ROV Dive Processing

SOI advances oceanographic research, discovery, and knowledge, and catalyzes sharing of information about the oceans. SOI collects a variety of ocean and atmospheric observational data from multiple instruments from their platform, the R/V Falkor. The clinic team developed a dashboard for the viewing of ocean video with important environmental measurements. They developed a data pipeline for multiple environmental measures from the ROV and computer vision models for tagging video frames.

3. Train the next generation of public interest data science leaders through hands-on experience.

The Clinic has provided students with the opportunity to challenge themselves with real-world data, clients, and problems while still having the support, mentorship, and resources of a classroom setting. Their ability to overcome common challenges in this course such as working with imperfect datasets, applying models and algorithms to real-world data, and navigating security and privacy issues has left them more prepared to tackle these problems upon graduation.

Additionally, students have gained invaluable practice in the realm of communicating results to a diverse set of stakeholders (e.g., industry, public interest, government agencies), and translating information into actionable insights, policy briefs and software prototypes. Each team, or group of students, is required to put together a video presentation at the end of each quarter outlining what they built and why it was impactful. These videos are shown in a quarterly “Clinic Showcase” which are open to the public and include faculty, staff and students from around campus. Teams of students also hold an open Q&A session after their video is shown for anyone with further questions.

To date 88 UChicago students have enrolled in the Civic Data and Technology Clinic and completed projects for social impact organizations. We have drawn from a wider range of degree programs around the university than expected. The students who have participated in Clinic have been enrolled in the following degree programs:

- Astrophysics
- Computational and Applied Mathematics
- Computer Science
- Economics

- Fundamentals: Issues/Texts
- MA - Computational Social Science
- Mathematics and Statistics
- Masters Program in Computer Science
- MBA
- MA - Public Policy
- MS - Computational Analysis and Public Policy
- Philosophy
- Physics
- Sociology
- Statistics

Challenges and Lessons Learned

As the Clinic has grown and scaled its impact since the pilot in 2020, we have overcome a number of challenges related to 1) designing repeatable processes, 2) providing more structured support for students, and 3) working within the complexity of a large research university.

Designing repeatable processes: One of the early lessons learned is that working within the 10-week quarter system means that the pace and turnover of students and projects needs to happen very quickly. As new teams of students join existing or new projects every ten weeks, the more that the Clinic team can create templates for gathering project information and scoping the quicker that teams can start to have an impact on Day 1 of the quarter. Additionally, leveraging systems like Slack, GitHub, and Google Drive for collaboration and getting roadblocks removed quickly have helped students avoid potential delays and keep the lines of communication open.

More structured support for students: One area that we continue to grow in is how we can support students with more formal training in addition to the hands-on experience of the course. Students have indicated that they would appreciate more background on software development best practices, tools, and systems as they work through them in real-time. We plan to implement more lectures and background materials as part of the Clinic in the years to come.

Working within the complexity of a large research university: Building formal partnerships within the university system has proven more complicated and taken longer than expected. We now know that in setting up working relationships with organizations we need to build in significant lead time for formal legal agreements to be put in place.

Copies of publications and media

“New Clinic Leverages Data Science for Social and Environmental Causes”

<https://datascience.uchicago.edu/news/new-clinic-leverages-data-science-for-social-and-environmental-causes/>

“Public Interest Technology Grant Funds and Expands CDAC/Harris Civic Data & Technology Clinic”

<https://datascience.uchicago.edu/news/public-interest-technology-grant-funds-and-expands-cdac-harris-civic-data-technology-clinic/>

“Civic Data and Technology Clinic Showcase – Winter 2021”

<https://datascience.uchicago.edu/events/civic-data-and-technology-clinic-showcase-winter-2021/>

UChicago’s Civic Data and Technology Clinic Showcase YouTube Channel:

<https://www.youtube.com/playlist?list=PL0IrlAluK93G8auHi7lkm6UqrUP91sNDZ>

Certification

All University of Chicago activities conducted with the Grant funds were and are consistent with charitable purposes as set forth in Section 501(c)(3) of the Internal Revenue Code, and University of Chicago complied with all provisions and restrictions contained in this Agreement, including, for example and without limitation, those provisions relating to lobbying and political activity.