**Data Engineering Contractor**

Crime Lab New York is seeking an experienced data engineer for limited-term contract work to design and build a data processing pipeline for a complex administrative dataset (more information about the dataset below). This pipeline will serve as the data infrastructure behind a multi-year scientific evaluation of a recent large-scale reform in criminal justice.

We’re looking for someone who will both listen closely to the needs of our principal investigators and data scientists but will also lead the design based on their experience and established best practice. Our current projects use a series of small R & Python scripts with makefiles to control the workflow. This design works well for smaller projects and simpler datasets but we’re interested in pursuing a more robust solution for larger and more complex projects. The ideal candidate will help us develop that vision in the context of this project.

**Responsibilities will include:**
- Build a data pipeline that cleans, merges, and reshapes a large number of raw data files into several clean and tidy datasets. This pipeline must support our data analysis code, which is written in Python or R.
- Develop clear documentation that will allow our team to extend and support the pipeline after hand-off. This includes a framework for how we would incorporate new datasets (from the same source systems) into this pipeline.
- Provide guidance and mentorship for data science team members on software and data engineering best practices.

**About the dataset**
The dataset is a collection of 30-40 tables from 8-10 transactional databases. Transactions related to a single entity can appear in all of these different databases, and while there’s usually a straightforward merging strategy, that is not always the case. There’s a significant amount of event-level duplication that needs to be resolved but there’s known business logic for much of that resolution.

**To Learn More or Apply**
If you’re interested in filling this limited-term role, please email Greg Stoddard gstoddard@uchicago.edu to learn more.

**About Crime Lab New York:**
Crime Lab New York (CLNY) is an independent, non-profit research center of the University of Chicago. We gather experts from around the nation, spanning disciplines, to develop data-driven approaches and generate rigorous, objective, scientific evidence.

---

1 This design and philosophy was heavily influenced by Patrick Ball’s “Principled Data Processing” (blog post: https://hrdag.org/2016/06/14/the-task-is-a-quantum-of-workflow/).

about what works in reducing crime and addressing critical problems facing society. We partner with city agencies, civic organizations and others to use data to identify, develop and evaluate interventions that can be measured and replicated. Our work employs innovations in natural experimentation, randomized controlled trials, behavioral economics, and machine learning. We use these tools to help policymakers identify and target interventions, make the best use of public resources, and benefit New York’s most disadvantaged citizens.

Crime Lab New York regularly seeks out passionate, data-oriented, public service-minded individuals to join our team. Crime Lab employees are committed to creating a collaborative, high energy office that is open and welcoming to individuals of all backgrounds and life experiences. We prioritize developing a workforce that is as diverse as the city in which we work.

For more information about the Crime Lab New York, go to [www.crimelabnewyork.org](http://www.crimelabnewyork.org).

All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, age, protected veteran status or status as an individual with disability. The University of Chicago is an Affirmative Action / Equal Opportunity / Disabled / Veterans Employer.

Job seekers in need of a reasonable accommodation to complete the application process may contact the Shared Services Office by calling 773-702-5800 or by emailing sharedservices@uchicago.edu with their request.