Urban Science Intensive II

Summer 2021

Final Report

Mapping Construction and Demolition Waste Flows for Assessing Recovery and Reuse Potential in NYC's Capital Program

JP McKay

Parth Singhal

Dina Wagdy

Table of Contents

Abstract

This report describes the process of visualizing the flow of construction and demolition waste (CDW) generated in the Long Island and New York City regions in order to provide context for policymaking that encourages local CDW recycling and reuse. It provides an overview of the transformation of multiple sources of regulatory reporting into a structured format, the amalgamation of the structured data into a single machine-readable dataset, and the production of a spatial visualization tool for interactive inquiry of the dataset by nontechnical users. The resulting product provides policymakers and industry stakeholders a means of spatially examining annual flow trends in aggregate by material types, transactions, and destinies (i.e., disposed, recycled, or designated as beneficial use). The dataset behind the tool enables other data professionals to perform analysis on historical CDW flows and thus better plan for the sourcing of locally produced CDW as an ecologically friendly material resource.

Introduction

New York is a city of ambition. In its 2015 OneNYC Plan, the mayor articulates planned actions to achieve the city's "80x50" commitment to reduce greenhouse gas (GHG) emissions eighty percent by 2050 and become the "most sustainable big city in the world." The plan comprises multiple sectors and services, including energy production, transportation, and solid waste management, calling for an ambitious reduction of waste disposal – ninety percent by 2030 relative to 2005 levels (The City of New York, 2015).

As of April 2021, active new construction and major alterations in New York accounted for over 175 million square feet of floor space (NYC Department of Building). This work comes with environmental costs, as many building material life cycles in the city today are linear, not circular. Projects often utilize materials from virgin sources, which are transferred long distances to the construction site and again at the end of their life to disposal. This contributes GHG emissions directly through fossil-fueled hauling and indirectly through the embodied carbon of the materials, present in the lost opportunities for their reuse. Worldwide, construction and the manufacturing of building materials accounts for eleven percent of GHG emissions (International Energy Agency, 2019).

Neither the city's existing climate legislation nor the OneNYC Plan address construction and demolition waste (CDW). Its major building-related climate law, Local Law 97, focuses on buildings' efficiencies, not their embodied carbon, and OneNYC's waste disposal goals only apply to putrescible waste. This is, in part, because CDW in New York is regulated at the state level. The New York State Department of Environmental Conservation (NYSDEC) regulates CDW and collects data on CDW from private waste haulers, transfer stations, recycling facilities, and landfills through mandated reporting requirements.

While city officials may not regulate CDW directly, they do hope to lead in inducing demand for CDW-derived materials by demonstrating its suitability and affordability as an alternative resource. As a start, the New York City Department of Design and Construction (NYCDDC), the primary client for this project, has expressed interest in incorporating CDW reuse provisions in contracts for its own capital projects. However, without understanding recent supply and demand for local CDW, the city cannot justify these provisions because it cannot estimate the financial feasibility of using CDW-derived materials.

In addition to supporting this direct goal of the NYCDDC, we posture that our tool will also enable visual exploration of CDW supply and demand by non-technical subject matter leaders elsewhere who are equipped to recognize patterns and trends that can inform the direction of further data analysis within their department or business. With this combination of exploratory and formal analysis, we believe organizational users will be able to cull sufficient evidence as to whether a CDW-related policy or investment can be justified (see *Appendix I* for examples of how the data may support specific policy options). To support this broader goal, we provide a framework for updating and integrating new data annually, assuring the tool's relevance continues (*see Appendix II*). One day, we hope the tool could even support the development of a formal, public-private marketplace for CDW-derived materials.

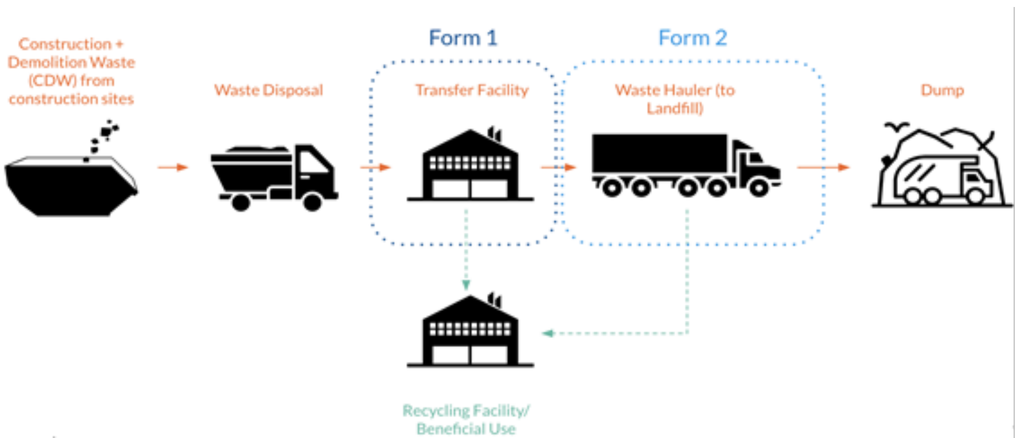

Figure 1: Typical CDW Flow in New York State (courtesy of NYCDDC)

Problem Statement

Our original hypothesis concerned itself with the inefficient distance CDW is often hauled, as we believed this held the greatest potential for our client's overarching aspiration to lower emissions attributable to CDW. As we learned from industry interviews and conversations with the NYCDDC, the New York State Department of Environmental Conservation, and the New Yok City Department of Environmental Protection, however, the emissions cost attributable to trip distance is 1.) likely less a factor in the overall emissions of a material's lifecycle than is the embodied carbon of the material, 2.) partly necessitated by New York City's strict regulation on in-city facilities, and 3.) a result of market forces more difficult to influence compared to the overall diversion rate of CDW.

While we still find the spatial aspect of our tool offers value to NYCDDC's main objective (i.e., understanding how, when, and where the city can better diversion rates for its own capital projects), the broader goal of our client to impact CDW diversion sector-wide lends itself to a focus on different variables of analyses. As such, our primary hypothesis is both more policy and diversion focused:

> *Wider visibility into diversion rates for specific material streams will encourage policymakers to support actions that increase CDW recycling by providing a scope of the opportunities present in the local recycled materials market.*

While this hypothesis cannot be proven before wider marketing of our application, it is measurable through the interest in the issue by policymakers with access to our tool and/or data. Based on our industry interviews, we believe increased market visibility will encourage political will to increase CDW diversion.

Literature Review

In the most recent systematic literature review of CDW research from Wu et al. (2019), the authors observe that, like our project, the most common methodology demonstrated for effectively modeling CDW flows relies on “widely-adopted bottom-up waste management data report[ing].” They go on to highlight the shortcoming of this approach, the difficulty coordinating consistent data reporting cross-regionally – a limitation we also face.

Jenny et al. (2016) articulate evidence-based design principles for spatial flow mapping that assisted us in augmenting our tool’s interpretability. These principles along with those identified by Yang et. Al (2019) are further explained in the Methodology.

It is worth noting too that our client’s vision of a circular construction industry in New York are not unheard of elsewhere. Japan has developed national legislation encouraging high-value CD&W waste reuse (Jin & Chen, 2015), while China has declared CDW reuse a national priority (Li et al., 2020). In western countries, examples of national frameworks exist, but much of the market innovation in CDW recycling and reuse comes from subnational and private sources. These include bottom-up policy proposals from the academic community in the United Kingdom (Ghaffar et al., 2020), a market platform for CDW as a commodity from a private startup based in the Netherlands (Excess Materials Exchange, n.d.), and the city of Austin, Texas’ 2016 construction and demolition recycling ordinance that requires permitted projects over 5000 sq ft. to divert at least fifty percent of CD debris produced (City of Austin, Texas, n.d.), among others.

Finally, while not features of this version of our product, we did explore options that could one day lend predictive and analytic capabilities to the tool. These include the algorithm proposed by Qiao et al. (2020), a robust prediction algorithm for CDW as a function of building area estimates and planned construction, as well as the approach of Guo and Zhu (2014) for detecting high-level patterns in spatial flow data.

Data

Before this project, no publicly accessible source aggregated information on CDW flows in New York. The only raw data on the topic exist as scanned copies of NYSDEC annual *Solid Waste Facility Forms*. As established by New York State's Environmental Conservation Law, NYS DEC regulates solid waste management facilities, requiring annual operating reports from permitted and registered facilities. Among other types of facilities, this includes CDW handling and recovery facilities, CDW-specific landfills, and other solid waste landfills that may collect CDW as part of their operations. Past year's reporting forms are available publicly by request to NYSDEC. A year's forms are typically available around April of the successive year.

These facility forms provide high-level data about CDW transport, processing, and disposal. Facilities may complete these forms digitally or by hand and submit them either digitally or via mail. As such, the text contained in the pdf versions of these forms range from partially computer recognizable to only human readable. NYS DEC had, however, organized pdf copies of submissions by year and reporting region in a standard hierarchal file directory.

Data Wrangling

Considering the range of computer readable formatting alongside the liberties taken by facility representatives in recording data (see *Appendix III: Examples of Optical Character Recognition Pitfalls in .pdf Forms* for examples), most of our dataset had to be derived manually. The heterogeneity of formatting was too wide to warrant custom scripting solutions for parsing and mapping data fields, but the sheer volume of form submissions across reporting years necessitated culling the scope of the entry process. As such, the dataset includes only areas of immediate interest to the client, not all regions in New York State (Table 1).

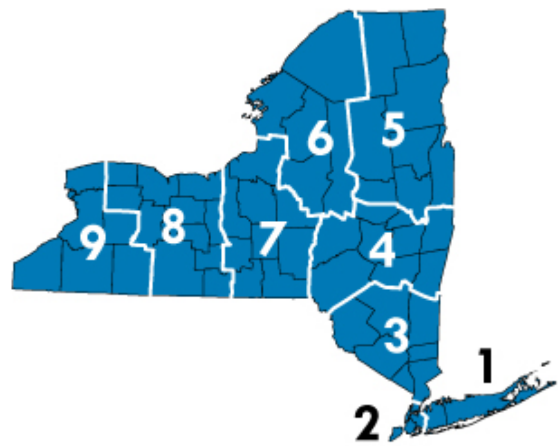

Figure 2: NYSDEC solid waste facility regions (courtesy of NYSDEC)

Table 1: Data collected by region, year, and facility type

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8
2020	Transfer & Landfill	Transfer & Landfill	Landfill only	Landfill only	Landfill only	Landfill only	Landfill only	Transfer & Landfill
2019	Transfer & Landfill	Transfer & Landfill	Landfill only	Landfill only	Landfill only	-	-	Transfer only

Manual data entry relies on the quality of information submitted. In some cases, facility representatives fail to provide detailed location addresses for service areas (i.e., sources or destinations) or provide total annual quantities rather than monthly quantities. With respect to the first scenario, we attribute the service areas of those trips to the respective town or county and make note in the visualization. For the latter case, we simply divided the total quantities by twelve; this issue did not occur frequently for large volumes and a more sophisticated approach would have cost significantly more time without reaping considerable benefit.

While material categories are provided, facility representatives also have the option of writing in their own. We aggregate these aliases as they are reported into twenty-one material categories. Determination of these categories was closely guided by the material categories provided on the form as well as the client's expertise (see ***Error! Reference source not found.***).

Data Cleaning

We undertook extensive data cleaning measures to ensure the data's integrity across all datasets, including manual confirmation of facilities with single-year or no entries, verification of numeric data suspected to contain error, and sanity checks of the aggregate data. Once cleaned, we transformed the raw entries, each consisting of unique facility, material, and source(s) (in the case of those facilities reporting incoming materials) or destination(s) (in the case of those reporting outgoing) groupings. From hence forth, the reported source or destination is generically referred to as a *service area*, while the facility is referred to simply as *reporting facility* (since materials are regularly transported between two facilities, a service area may be a facility in one entry and the reporting facility in another). Once this initial dataset was created, with source(s) or destination(s) associated with each reporting facility-material grouping, a custom Python script further split each entry into reporting facility, material, and *single* source/destination groupings. In this penultimate dataset, sources and destinations are both found in the same column, with a separate column indicating the flow direction relative to the reporting facility ("incoming" or "outgoing"). The resulting rows each correlate with a volume of a material moved between two nodes, the direction of movement, and identifying locational information of each node.

Geocoding

The final step to prepare the data was to connect every reporting facility and service areas to mappable coordinates, a process known as geocoding. Using GeoPy, a geocoding Python client, and GoogleV3, a Google Maps API, we mapped each reporting facility and each service area to its respective coordinates. The result is a georeferenced network of directional, material specific source-destination trips that can be overlain on any arbitrary map interface, as is done in the application.

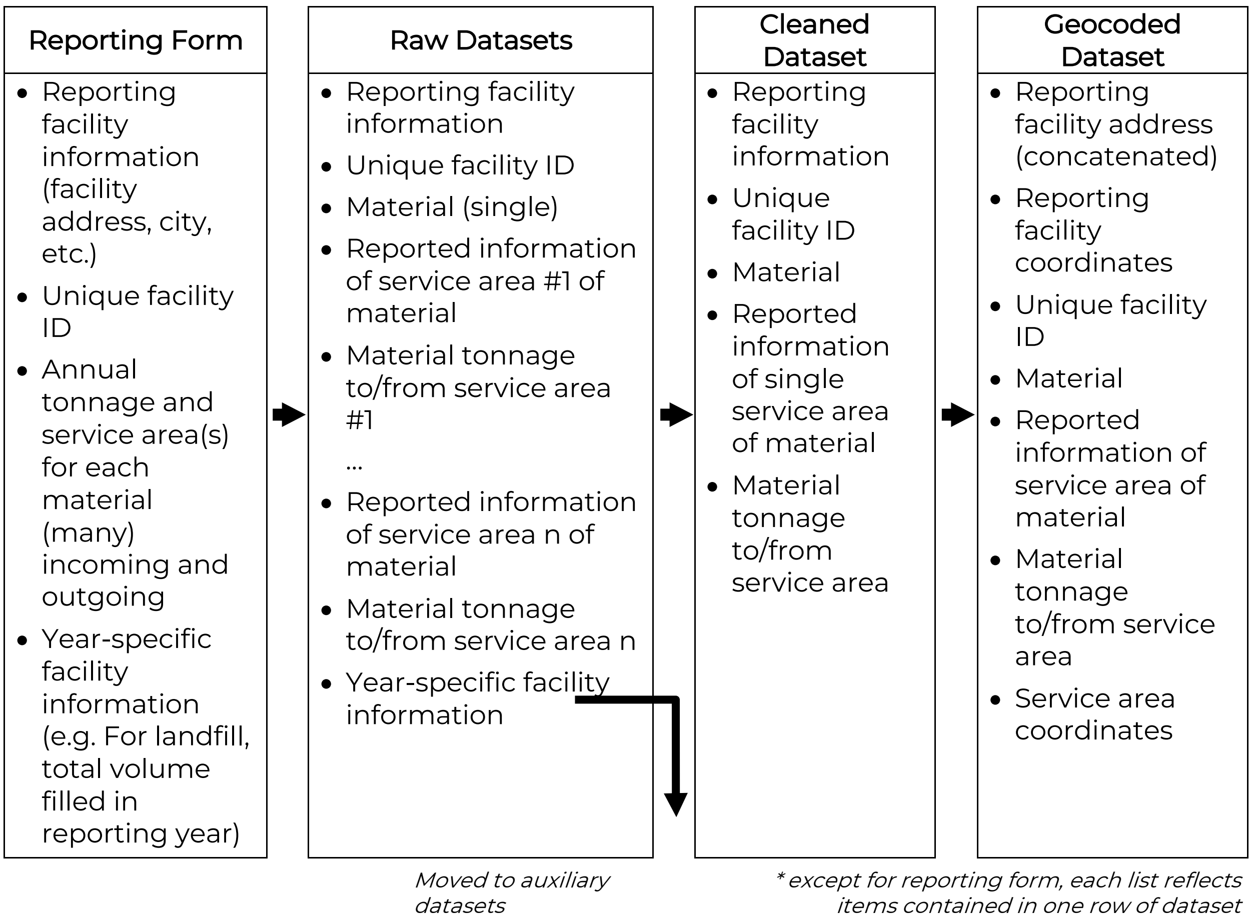

*Figure 3: Transformations to Final Geocoded Dataset**

For reporting facilities, we confirmed each facility address as listed by the facility representative by referencing it to the address listed for the facility in NYSDEC's [Solid](#)

Waste Management Facilities dataset. Service areas, on the other hand, could be not only a solid waste management facility, but any location, in or out of state, where CDW was generated. Whether a facility or arbitrary address, most facility representatives reported service areas colloquially rather than with a formal address (see Table 2 for an example). This was problematic for two reasons: 1.) the geocoding script might have wrongly identified these entries as many different locations rather than attribute the trip entries to the one service area, and 2.) the lack of detail may have made it difficult for GoogleV3 to accurately geocode these entries. In this case, we cross-referenced and corrected all service areas to ensure they had uniform details so that they would be correctly geocoded.

Table 2: Example of address needed for geocoding service area compared to actual entries

Needed Facility Name & Info. For Accurate Geocoding	Examples from Filled Forms
'110 Sand Company, Melville, NY'	110 Sand
	110 Sand Co.
	110 Sand Company
	110 Sand and Gravel

The final dataset is robust at the trip level, offering material-category-specific volumes by source and destination. Limitations of the data are outlaid in detail in *Appendix V – Data Limitations.*

Methodology

Before building the application, we surveyed several leading Python data visualization libraries. We required a library that would allow interactivity; provide a module or modules supporting flow mapping, ideally built with best practices in mind for flow map interpretation; and integrate with a high-level, Python-centric web framework. We ultimately chose PyDeck, a set of Python bindings for the JavaScript spatial visualization rendering package Deck.gl, for our visualization library. We also use Plotly to render supplementary, non-map figures. We use the Streamlit API to provide a straightforward web interface for interacting with the application.

Plotly provides an extensive, interactive graphing library, and was originally our choice for all visualizations accessible through the application. However, when weighing it against PyDeck for the map visualization, the main feature of our application, we found PyDeck a more modern and interpretable alternative. This comes, in part, from its reliance on DeckGL, designed by Uber Inc. specifically for visualizing large-scale spatial datasets. We found the prebuilt layer most akin to our use case ("ArcLayer") addressed many of our aesthetic concerns to a degree that Plotly's did not, primarily its automatic generation of curved, symmetric flow lines, one of the major design principles recommended by Jenny et al. (p. 65) for flow mapping. Finally, because DeckGL is in turn built on the popular three-dimensional graphics API WebGL, the map renders in a 3d environment. Research on three-dimensional flow mapping for virtual and augmented reality cases suggests users consistently prefer 3D visualization for the ease it lends in interpreting flows, especially where arc height correlates with distance between origin and destination (Yang et al., 2019). We find this preference remains when viewing flow maps in an on-screen 3D environment, further supporting our choice of PyDeck for map visualization.

When considering libraries to abstract front-end development into something suitable for this project, we found Dash and Streamlit to offer similarly intuitive features. Besides being compatible with both Plotly and PyDeck, the two libraries similarly allow the creation of layout components through simple. declarative statements as well as abstract all JavaScript, CSS, and HTML elements, necessary features for a team like us with little web experience. We ultimately chose to use Streamlit because the company behind it offers free hosting for non-enterprise users, whereas Dash only offered a paid option marketed towards enterprise-level projects.

Results

As mentioned in *Problem Statement*, we aim to provide wider visibility into the flows of CDW in New York City. Also, as we highly prioritize the sustainability, relevance, and reliability of our data, we chose to capture our findings, both quantitative and qualitative, and to represent them via the following media:

- The web application hosting an interactive map of CDW trips and other statistical visualizations
- The associated datasets
- A list of recommendations for future research (*Recommendations for* Future Research)
- Recommendations for updating and automating NYSDEC's current method of reporting (***Error! Reference source not found.***)

Web Application

The web application provides the user a dynamic visualization of CDW flows for the New York City and Long Island regions. For instance, if one were interested in studying the flow of general C&D debris through transfer facilities in 2020, the webapp would produce a color-coded arc for each trip describing the direction of the flow (Figure 4). When the user hovers over the arc, they can see the details of the trip in terms of material, source, destination, weight in tons and destiny, when applicable. The destiny is only applicable to "outgoing" trips, as they indicate whether materials were transferred, recycled, disposed, or used as alternative daily/operating cover when leaving transfer facilities or if they were recycled, disposed, or used as alternative daily/operating cover at landfills.

Visualizing these flows allows the user to understand where the waste is generated from and typically, how far it travels. It also allows them to compare the flows between different time periods. Figure 5 shows flows of general construction & demolition debris to and from transfer facilities in 2019, where one can clearly see similarities and differences between the years' flows. In both scenarios, general construction & demolition debris tends to originate primarily from the NYC and Long Island regions and is disposed in farther areas of New York State such as Seneca County, Ontario County, and Albany County, as well as other states. The main difference, however, is that in 2019, the debris travelled to farther states such as Ohio, Arkansas, and Virginia, compared to just the tristate area in 2020. Additionally, there was a significant decrease in the volume of waste hauled and number of trips in 2020 compared to 2019, likely attributable to the economic effects of the pandemic.

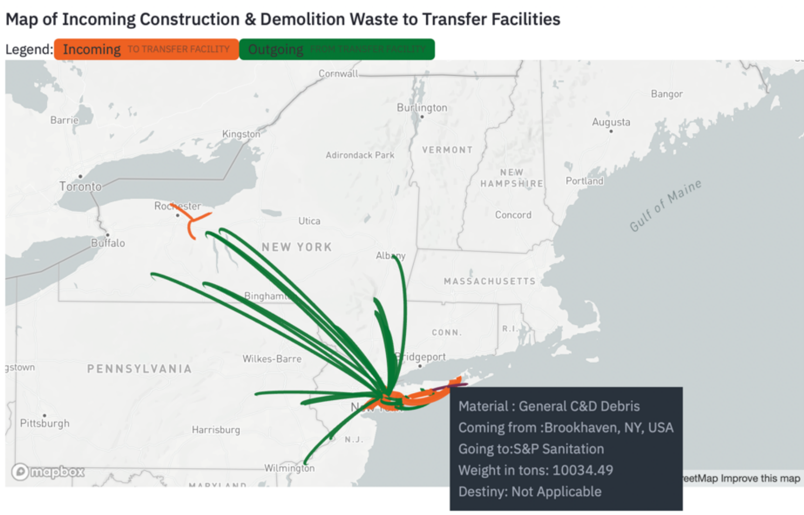

Figure 4: Snapshot from webapp: flow of general construction & demolition debris to and from

Figure 5: Snapshot from webapp: flow of general construction & demolition debris to and from

Similar visualizations can be created, as the user has full control of the map. A user can zoom in, zoom out, pan, and rotate the map as they please. Furthermore, similar conclusions drawn for any of the twenty-one materials groups.

Other Visualizations

In an effort to increase the visibility of our data, we embedded a Sankey diagram tool in the webapp that breaks down the material that reaches the facilities and visualizes that breakdown in terms of volume and destiny. Figure 6 shows the breakdown of materials that reached region 4 Landfills in 2020. As shown, almost all materials were used as cover at the landfills except for a small sliver of around 30 tons of Metals and Tires, which were recycled. The webapp can provide this breakdown for other regions

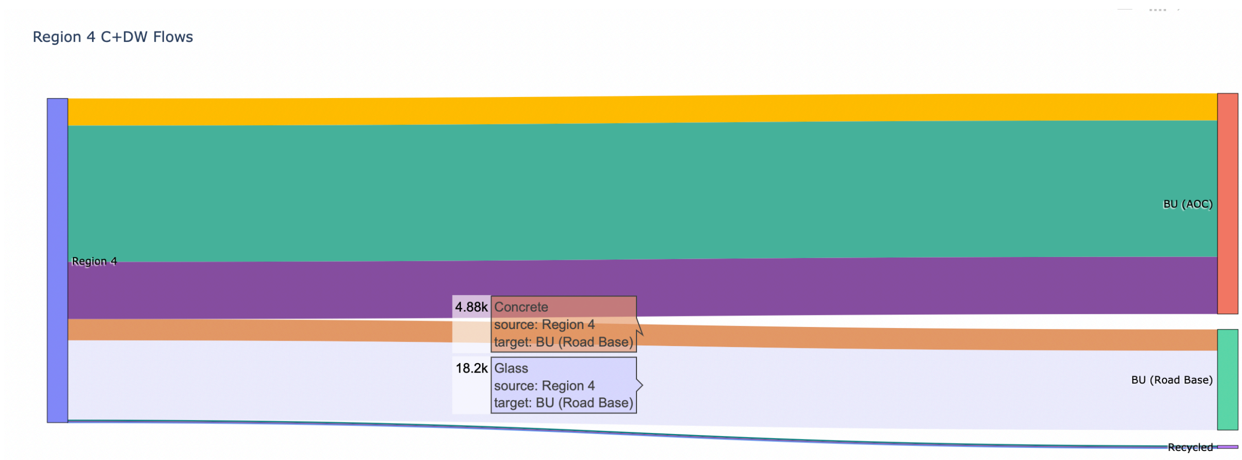

Figure 6: Breakdown of materials reaching region 4 landfills in 2020

as well as for transfer facilities.

Lastly, we embed a stacked bar plot in the visualization tool to show the monthly breakdown of generated waste each year. This represents the total waste that was reported by facilities each year. Figure 7 shows the breakdown of material in 2020. One can clearly see a significant decrease in overall reported waste in April, likely as a result of the pandemic, while June shows the highest rate of incoming waste, possibly due to increased construction activity to make up for the drop that occurred in April and May. As a basis for comparison, Figure 8 shows the monthly breakdown of generated CDW in 2019. There are a few clear differences between the years; for one, there's a clear gradual increase in generated CDW that peaks in October in 2019, unlike 2020. Also, the proportion of materials seem to be quite consistent in each month of 2019 compared to 2020. Finally, although we've collected data for more regions in 2020 than 2019, the reported volume of CDW in 2019 is 3-4 times larger than that reported in 2020. However, the main similarity between both years is that the most voluminous categories of generated CDW are "General Construction & Demolition Debris," "Concrete," and "Fill."

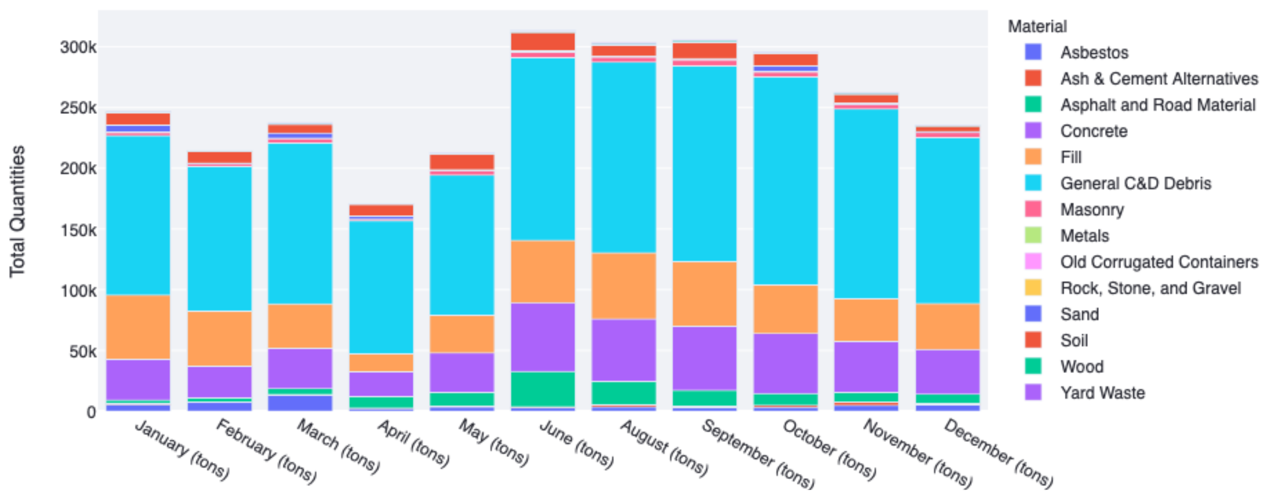

Figure 8: Monthly generated CDW material breakdown in 2020

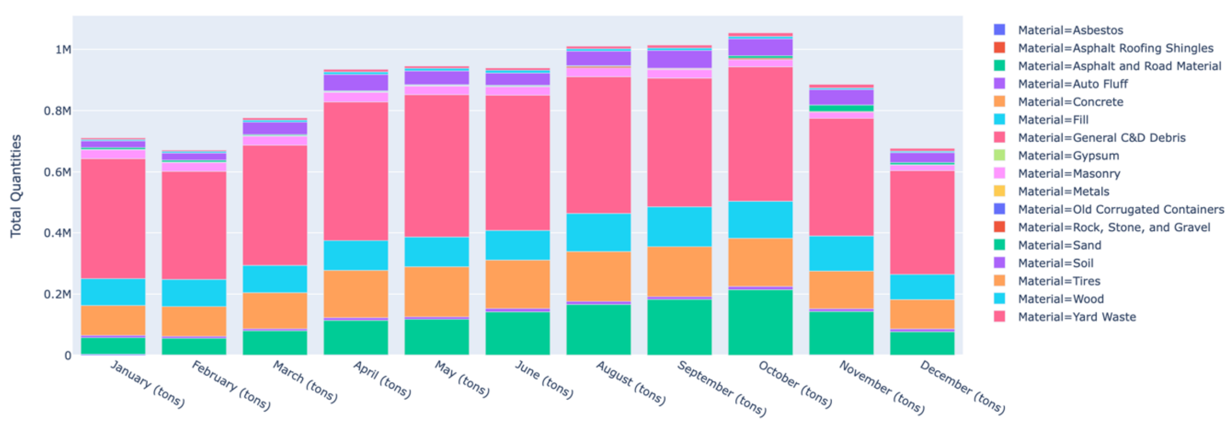

Figure 7: Monthly generated CDW material breakdown in 2019

Put together, these tools provide users with a mean to visualize how much CDW is generated, where it comes from, where it goes, and what its final destiny is when it reaches its destination. This sets the groundwork for deciding what and how policies can be drafted and enacted to have the most impact on CDW and effectively tackle the consequences that result from its generation.

Conclusion

This project represents but one stone in the foundation of an expansive, actionable knowledge repository for the built environment being built at NYCDDC. NYCDDC's *Town+Gown*, a city-wide university-community partnership program and the primary client for this project, as well as its expert-comprised Urban Resource Recovery Group, have taken many steps since 2016 supporting engineering, policy, and legal research to innovate how New York approaches CDW reuse and recycling. We are thankful for Terri Matthews, Director of Town+ Gown, for her persistent support during this four-month project, as well as for the expertise of Jennifer McDonnell at the New Yok City Department of Environmental Protection, Kathy Prather at the NYSDEC, and all the industry experts we had the privilege to learn from. We look forward to following how the application, dataset, and supplementary deliverables integrate with and influence this growing body of timely research.

References

Climate Mobilization Act. (n.d.). New York City Council. https://council.nyc.gov/data/green/

Excess Materials Exchange. (n.d.). *About Us – Excess Materials Exchange*. Retrieved May 2, 2021, from https://excessmaterialsexchange.com/en_us/about-us/

Ghaffar, S. H., Burman, M., & Braimah, N. (2020). Pathways to circular construction: An integrated management of construction and demolition waste for resource recovery. *Journal of Cleaner Production*, *244*, 118710. https://doi.org/10.1016/j.jclepro.2019.118710

Guo, D., & Zhu, X. (2014). Origin-Destination Flow Data Smoothing and Mapping. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2043–2052. https://doi.org/10.1109/tvcg.2014.2346271

International Energy Agency. (2019, December). *Global Status Report for Buildings and Construction 2019.* https://www.iea.org/reports/global-status-report-for-buildings-and-construction-2019

Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E., & Jenny, H. (2016). Design principles for origin-destination flow maps. *Cartography and Geographic Information Science*, *45*(1), 62–75. https://doi.org/10.1080/15230406.2016.1262280

Jin, R., & Chen, Q. (2015). Investigation of Concrete Recycling in the U.S. Construction Industry. *Procedia Engineering*, *118*, 894–901. https://doi.org/10.1016/j.proeng.2015.08.528

Li, J., Yao, Y., Zuo, J., & Li, J. (2020). Key policies to the development of construction and demolition waste recycling industry in China. *Waste Management*, *108*, 137–143. https://doi.org/10.1016/j.wasman.2020.04.016

NYC Department of Building. (2021). *Active Major Construction*. nyc.gov. https://www1.nyc.gov/assets/buildings/html/nyc-active-major-construction.html

Qiao, L., Liu, D., Yuan, X., Wang, Q., & Ma, Q. (2020). Generation and Prediction of Construction and Demolition Waste Using Exponential Smoothing Method: A Case Study of Shandong Province, China. *Sustainability*, *12*(12), 5094. https://doi.org/10.3390/su12125094

Slowey, K. (2021, March 3). *New York City to restart $17B of capital construction projects*. Construction Dive. https://www.constructiondive.com/news/new-york-city-to-restart-17b-of-capital-construction-projects/596029/

The City of New York. (2015, April). *One New York: The Plan for a Strong and Just City*. http://www.nyc.gov/html/onenyc/downloads/pdf/publications/OneNYC.pdf

Wu, H., Zuo, J., Zillante, G., Wang, J., & Yuan, H. (2019). Status quo and future directions of construction and demolition waste research: A critical review. *Journal of Cleaner Production*, *240*, 118163. https://doi.org/10.1016/j.jclepro.2019.118163

Yang, Y., Dwyer, T., Jenny, B., Marriott, K., Cordeil, M., & Chen, H. (2019). Origin-Destination Flow Maps in Immersive Environments. IEEE Transactions on Visualization and Computer Graphics, 25(1), 693–703. https://doi.org/10.1109/tvcg.2018.2865192

Appendices

Appendix I: Qualitative Findings from Interviews with Industry Experts

As part of the discovery phase, we conducted several interviews with New York-based experts in construction and demolition waste management to better understand how the application might be applied to support different policy proposals. Our interviewees included:

- Kendall Christiansen, former director of the local trade association representing haulers and recyclers and a private consultant with over 30 decades of experience in the municipal waste industry,
- Miriam Voss, Sustainable Design Coordinator for the Port Authority of NY and NJ, a quasi-public organization and partner of the city of New York, who is looking to do for the Port Authority's CDW management footprint what our client seeks to do for the NYCDDC, and
- Amanda Kaminsky, former architect and founder and principal of Building Product Ecosystems, a consultancy helping building owners innovate material resource cycles.

Mr. Christiansen provided background on the retraction of New York City's recycling and recovery facilities, the reason for which he cited as a gradual tightening of regulations in reaction to population densification beginning in the mid 1990s. As residential areas spilled over into what were once industrial districts, the city set requirements for facilities to better integrate their practices with new development. This included restrictions on the length of onsite storage of hauled materials, on diesel emissions, and on trucking activity. Along with rising industrial and commercial real estate costs and low transportation costs relative to processing costs, these regulations have contributed to the shrinking of CDW transfer and processing facilities within municipal borders.

Ms. Voss offered insight into the responsibility of project owners and managers in material management and how the design process integrates with the responsibility. She said that designers often do not recognize the opportunities for the use of recycled materials simply because they are not familiar or because it is not required by the client. With awareness and the proper integration of provisions in the contract before the bidding process, she believes contractors could regularly diverge 75-95% of materials from the waste stream. She also cited slow-to-evolve specifications, many of which she said are overly stringent and may limit the applications of otherwise suitable recycled materials. Finally, she advocated an alternate approach to lifecycle management itself, in which buildings are deconstructed rather than demolished at the end of their useful life.

Lastly, our conversation with Ms. Kaminsky centered on connecting city- and region-level data with meaningful action. She is of the opinion that normalizing and promoting national recycling certifications in the region would lead to a more accountable recycling industry, adding that the current dearth of certified recyclers in New York does not allow true price competition for certification-specified contracts. She emphasized the need for clear communication at early stages in the bid process between contractor and project management, both so that management understands a contractor's technical requirements for meeting a given diversion goal and so that contractors can accurately understand the intended lifecycle goals of the contract before bidding. Facility-specific capabilities rely on differing degrees of onsite source separation, a responsibility of the project.

All three experts also provided additional recommendations on material-specific challenges, suggestions for improved material management, and areas of knowledge gaps. While this information is outside the scope of our project, we intend to note these areas for improvement in our accompanying recommendations to NYCDDC.

Below, we provide an example of a policy lever suggested by one of our interviewees that might be informed by our application in order to increase the recycling rate for CDW. Brief examples of other policy levers, as suggested by experts, are summarized in Table 3.

Policy Lever Example – Increasing Certification

Through our interviews and follow-on research, we learned about the impact of certification on CDW recycling and about the current level of maturation of these certifications in the United States. The Sacramento-based *Recycling Certification Institute* is the dominant and most widely recognized certification body for CDW recycling in the United States. Since 2014, their certification has provided a protocol for verifying CDW diversion rates, which includes an independent, on-the-ground audit to confirm diversion rates are correctly calculated. (It is important to note that a certification does not specify a capability of a contractor to recycle waste, but rather that they describe their recycling capabilities in a consistent, uniform manner.)

The importance of certification, according to our interview sources, is to create standard and unambiguous language for diversion provisions in contracts. As an example, *alternative daily cover (ADC)* – a term for a mix of inert materials (e.g. crushed concrete, stone) often derived from demolition and excavation used to "cap" a landfill at the end of each day – may be thought of as "recycling" by contractors, who may include it in their diversion rate. In contrast, the New York Department of Environment and Conversation does *not* consider such use as recycling. Widespread certification would ensure that contractors are considering ADC under a definition congruent with that of the state. This is only one example of the merit of certification, which among other things defines "recycled" in detailed and material-specific language.

While Recycling Certification Institute's (RCI) certification is held by nearly two dozen contractors across the nation, most are in western United States, with only two RCI-certified facilities in the New York City region. If New York City were to craft an ordinance to promote certification, it may use data from our application to support the legislation by illustrating the common imbalance in material reported as recycled versus ADC.

Table 3: Examples of how evidence from application may support policy levers

Policy proposal	Suggested by	How our app may help inform
Require performance bonds for diversion rates to ensure contractors meet target rates	Kendall Christensen	Informing baseline diversion standards to set feasible bond thresholds
Revaluate landfill tipping fees to reflect cost of future negative externalities of disposal (ex. Excess toxic hydrogen sulfide gas from gypsum wallboard, which must later be remediated)	Amanda Kaminsky	Averaging tipping fees by location, tracking gypsum disposal relative to proximity of local population
Factor carbon impact of material into disposal costs	Amanda Kaminsky	Providing carbon estimate for specific project based on embodied carbon of material + weigh-distance hauled
Mandate implementation of Envision or LEED standards for certain types of contracts	Miriam Voss	Informing baseline diversion standard to exceed
Relax stringent material specifications based on intended use of recycled material (ex. allow use of lightly contaminated soil for fill under sidewalk rather than hauling residential quality soil from hundreds of miles away)	Miriam Voss	Compare concentrations of lightly contaminated material to concentrations of proposed appropriate uses

Appendix II – Guideline to Automate DEC Reporting Methodology

Based on our data wrangling, collection and creation experience, this brief guideline outlines changes that can be made to the current method of reporting that would optimize and simplify the process for both the DEC and the form-fillers at the facilities. This guideline prioritizes limiting human error as much as possible, ensuring relevant data is available in a timely manne,r and keeping the developed webapp relevant and reliable for future use. The guidelines does not wish to reinvent the wheel; in fact, it takes advantage of the DEC's available resources such as their fillable pdf forms and the existing Solid Waste Management Facilities (SWMF) Dataset, which holds vital information on each facility and was used extensively in creating our three datasets.

1. Create a simple portal to be hosted on the DEC's website that would allow facilities to register using their unique Activity Numbers.
    a. Based on each facility's Activity Description, the correct form(s) would be automatically available for filling.
2. Rather than giving the form-filler the liberty of typing in and completing all fields themselves, they would be limited to a set of pre-set fields to decrease the chances of human error and ensuring that the data provided is as useful as possible. For example:
    a. The first page of each form would be pre-filled, as it would be linked to the existing SWMF dataset, which already holds all the pertinent information on the facility.
    b. The list of materials should also be predetermined, allowing form fillers to choose from that predefined list to ensure proper categorization of the material and to avoid having thousands of variations of material names as shown in Appendix IV.
    c. Service areas can again be linked to the SWMF dataset, since, in most cases, the material is typically received from or sent to other facilities that exist in that dataset. This ensures that their correct name, address, facility ID and geolocation are recorded and avoids duplicate reporting of the

same trip (refer to the last point in Appendix V). It also means that the service area will automatically be geocoded allowing it to be mapped on the interactive tool.

   d. In cases where the service area does not exist in the form and a unique address needs to be provided, all necessary fields required to properly geocode the new service area can be requested and an error could pop-up if they're not filled properly/completely. This resolves the main data cleaning issue that we faced and the second point in data limitations (Appendix V).

3. Each field of the pdf form can be hyperlinked to its relevant field in the csv file. In that case, once a form is submitted and if the csv files are hosted on the web, the dataset will be automatically updated with the new data.
4. With the datasets being hosted online, the code can be adapted to pull the data using an API and automatically generate a new arc displaying each trip as soon as the form is submitted.

After studying the forms extensively, we believe that contractors would welcome and appreciate this updated method of reporting especially since we noticed that they often run out of space when completing the form and resort to providing supplementary attachments (see examples from Waste Management Company and Seneca Meadows Landfill in region 8). The portal could provide an option to add more pages or rows to accommodate this.

Appendix III: Examples of Optical Character Recognition Pitfalls in .pdf Forms

SERVICE AREA OF SOLID WASTE RECEIVED (where the waste is coming from)					
TYPE OF WASTE	SOLID WASTE MANAGEMENT FACILITY FROM WHICH IT WAS RECEIVED (Name & Address) OR *"Direct Haul"*	SERVICE AREA STATE OR COUNTRY	SERVICE AREA COUNTY OR PROVINCE	NYS PLANNING UNIT (See Attached List of NYS Planning Units)	TONS RECEIVED
Asphalt Millings	Direct Haul	N.Y.	Albany	CRSWMP Colonie	
Asphalt Pavement					7,372
Asphalt Roofing Shingles					
Brick					2,000

A reader would try to capture the arrows and x's which would make it difficult to accurately read and record the items in cells.

Construction & Demolition (CD) Debris	Weight ~~(tons)~~ YDS 3
January	
February	150
March	
April	
May	150
June	
July	
August	150
September	
October	
November	225
December	
Total Disposed For Year	675
Daily Average (Tons)	

In this case, the reader would accurately capture the numbers, however wrongly associate it with one month instead of three. Also, the units have been changed to cubic yards from tons which wouldn't be captured by the reader.

Appendix IV – Categories of waste materials based on reported aliases

Category	Reported Material Aliases
Concrete	Concrete', 'Concrete/Stone', 'Aggregate + Concrete', 'Aggregate/Concrete', 'Concrete/Dirt', 'Concrete Blend', 'RCA', 'Recycled Concrete Aggregate'
Asphalt & Road Material	Asphalt and Road Material', 'Asphalt', 'Asphalt Millings', 'Asphalt Pavement', 'Road Material', 'Road Building Material', 'Street Excavation Material', 'Concrete-Asphalt Mix', 'Mixed Broken Asphalt', 'RAP', 'Rap'
Metals	'Metals', 'Nonferrous Metal', 'Aluminium', 'Bulk Metal', 'Bulk Metal (from C&D Debris)', 'Bulk Metal (from CD debris)', 'Bullk Metal', 'Bullk Metal (from C&D Debris)','Bulk Metal (from MSW)', 'Bulk Metal/Appliances', 'Other Metal', 'Metal', 'Mixed Metal', 'Electronics', 'Scrap Metal', 'Wire'
Masonry	Masonry', 'Brick', 'Other Masonry Materials', 'Other Masonery Materials
General C&D Debris	General C&D Debris', 'Construction & Demolition Debris - out of county', 'Construction & Demolition (C&D) Debris', 'Construction and Demolition Debris', 'Construction & Demolition Debris', 'Mixed Loads', 'Mixed Debris', 'Mixed Load', 'Unprocessed Material', 'Unprocessed Mix', 'Brick, concrete','C&D Debris', 'Mixed - no Asphalt', 'Mixed Concrete + Brick', 'C&D Blend', 'Processed C&D','Mixed Concrete and Brick', 'Mixed Mixture of Asphalt Brick + Dirt'
Rock, Stone & Gravel	Rock, Stone, and Gravel','Rock', 'Rock/Stone', 'Stone', 'Gravel', 'Blend/Stone', 'Gravel/Rock/Stone', 'Rock and Dirt', 'Blend/Stone', 'Drainage Stone', 'Clay, Rock
Soil	'Soil', 'Topsoil'
Wood	Wood', 'Unadulterated Wood', 'Waste Wood', 'Wood', 'Woodchips', 'Wood/Wood Chips', 'Pallet Wood Chips', 'Unadulterated Wood (chips)
Old Corrugated Containers	'Old Corrugated Containers', 'OCC', 'Cardboard', 'Commingled Paper (all grades)','Corrugated Containers', 'Commingled Containers', 'Commingled Paper', 'Corrugated Cardboard', 'Confidential Paper', 'Mixed Paper', 'Mixed Containers', 'Paper Mill Sludge', 'Short Paper Fiber','Paper/Cardboard'
Fill	'Fill', 'Bulk', 'ADCM', 'Restricted - Use Fill', 'Restricted-Use Fill', 'Processed Fill', 'Screened Fill', 'Screening', 'Mixed Fill', 'Concrete, Brick, Dirt', 'General Fill', 'General', 'Excavated Mix: rock, conncrete, asphalt', 'Mixed Fill', 'Limited-Use Fill', 'Screen Fill 1/4"', 'Structural Fill', 'Unprocessed Fill', 'Soil/Sand', 'Concrete, Stone, and Screening','50% Dirt', 'Bentonite', 'Clay', 'Excavated

	Material', 'Clean Fill', 'Contaminated Soil','Dirt','Dirt/Concrete/Mixed Agg', 'Trench Fill', 'Blasting Sand as AOC', 'Bulk', 'Construction & Demolition (Cover)', 'Mixed Concrete, Asphalt, Dirt + Brick', 'Mixed Concrete, Brick, Dirt', 'DPW/Animal Shelter ADCM', 'GNF Cover ADCM', 'Special Waste (Cover)', 'Special Waste/Sludge/Other (Cover)', 'Grit', 'In house AOC blend', 'Industrial Waste (Alum AOC)', 'Petroleum Contaminated Soil', 'Processed C&D mix with sludge', 'Sandblast Grit', 'Solidification Bulking Agent (SBA)', 'Special Waste/Sludge/Other (Cover)', 'Coal, Sand, Rock','Petroleum Contaminated Soil - out-of-county'
Asphalt Roofing Shingles	'Asphalt Roofing Shingles', 'Roofing', 'Asphalt Roof Shingles'
Sand	'Sand', 'Sand (Bankrun)', 'Foundry Sand
Residue	Residue
Tires	'Tires', 'Waste Tire-Derived Aggregate', 'Waste Tires
Other	'Other', 'Commingled Containers (metal, glass, plastic)', 'Plastic', 'Batteries & Bulbs', 'Brown Stock', 'Cortland County Recycling Variance', 'Treated Regulated Medical Waste', Punch Outs', 'Incinerated Sewer Sludge', 'MSW', 'MSW Ash','Miscellaneous','Mixed Municipal Solid Waste', 'Other', 'Sewage Treatment Plant Sludge', 'Single Stream', 'Textiles', 'Water Jet Garnet', 'Industrial Waste', 'Mixed Rigid Plastic', 'Mixed Municipal Solid Waste - out-of-county','Sewage Treatment Plant Sludge - out-of-county'
Gypsum	Gypsum
Yard Waste	Yard Waste', 'Brush', 'Brush, branches, stumps, logs, wood chips','Bush, Branches, Trees, Stumps', 'Leaves', 'Yard Waste', 'Tree Debris', 'Brush, Branches, Trees & Stumps', 'Bush, Branches, Trees & Stumps', 'Compost', 'WWTP to compost', 'Land Clearing Debris', 'Landfill - land clearing debris','Mulch','Colored Mulch','Stumps & Branches', 'Compost & MRF Residuals'
Asbestos	Asbestos
Ash & Cement Alternatives	Ash & Cement Alternatives', 'Ash', 'Ash (MSW Energy Recovery)', 'Wood Ash', 'Industrial Ash', 'Industrial Waste (Alum Sludge Powder)
Glass	Glass
Auto Fluff	Auto Fluff', 'Auto Fluff (Cover)', 'Shredder Fluff', 'Car Fluff'

Appendix V – Data Limitations

Data for this project have been derived from publicly available operating reports for permitted and registered solid waste management facilities, primarily from Part 360 reporting regions 1, 2, and 8, as they were submitted to the New York State Department of Environmental Conservation for 2019 and 2020. As described in *Data*, the integrity data have been verified through multiple measures. However, there are a few limitations to this data that should be noted depending on the application:

- Reporting facilities are not required to state the unique "Facility ID" of their service areas or the date in which a trip took place, rendering it virtually impossible to link trip data together. For example, if facility A reported that fifty tons of concrete were transferred to facility B and facility B reported receiving fifty tons of concrete from facility A, the method by which those two reports are filled and submitted does not allow us to link those two trips together as one. The forms could be submitted during two different years (in which case, they could be combined with other trips from the same facility and no longer reflect the raw fifty tons), the generic facility names are not easily linked to their respective IDs, or simply the facility name could not be provided. This is a limitation since trips will technically be double-counted and the user will have to decide which sets of facilities to trust to have clearer information.
- In some cases, total tonnage amounts were only collected for what can be strictly considered construction and demolition waste; hence, this data does not accurately encompass all volumes of other types of waste, even if it was reported as incoming or outgoing by a facility. This includes but is not limited to generic plastic products; yard debris; cardboard, paper, and other corrugated materials; reused furniture and architectural components; and unspecified municipal waste.
- While NYSDEC requests in-state origin and destinations to be tied to each material type at no higher than the county level, some facility representatives

do not always follow this convention and may report place names that are ambiguous. When they could be confidently linked to a geography, we have done so at the lowest reasonable level; otherwise, these entries were dropped. The few entries attributed as “NYC,” “New York City,” or similar were instead attributed to New York County (Manhattan) only, though future work may choose to impute these missing values according to the known distribution of each material type’s volumes across the five boroughs.

- As the data relies on human reporting, there’s no way of knowing that what has been reported fully reflects reality. We learnt from our interviews that data collected from RCI-certified facilities are significantly more reliable compared to their non-certified counterparts. With only two certified facilities in New York State and mixed adoption of robust reporting methods (e.g. sensors, scales, cross-checking reported tonnage at both ends of the trip), we can only trust that data that has been reported is accurate.

Appendix VI – Recommendations for Future Research

This section describes further research we would have liked to carry out had time permitted, as well as how we believe our data can provide insight to existing research about how ongoing construction activity in the city affects CDW generation. The following research topics could benefit from the novel datasets that we created:

- *Predicting the volume and rate of CDW based on construction activity from capital projects in NYC*. NYC Open Data hosts datasets on current and historical construction projects in the city; by connecting those datasets with the CDW dataset, we could create prediction models based on type, size and cost of capital projects to estimate the volume and rate of debris generated and sent to Transfer Facilities & Landfills. This is useful in estimating the ecological and environmental impact of a project, which allows decision-makers to reassess the materials they intend on using in their projects or even set limits for contractors on how much waste can be generated from their site, incentivizing them to reuse and recycle potential CDW when possible.

- *Building a CDW marketplace*. The map provides a fundamental starting point if the city wished to implement a public-private marketplace for CDW. It provides rough estimates on how much waste is generated, where it is generated, and when; with this information, owners and contractors could better consider one another as potential vendors for CDW-derived materials, especially if they are more closely located than virgin material sources.

- *Studying the effects of CDW generation on air quality*. Again, the NYC Open Data website holds information on the air quality in NYC from various air quality sensors that collect data throughout the year. Additionally, other privately-owned but publicly available air quality datasets are available and can be spatially connected to the CDW trip data now that trips are geolocated. Studying the spatial correlation between air quality levels near landfill sites and transfer facilities and the rates by which CDW is generated and received by these facilities could shed some light on whether air quality is affected by CDW or not. In cases where it is, this may nudge policymakers and agencies to adjust limits on the type of materials and quantities that can be accepted by a facility.

- *Simulating truck routes based on trip data.* Such routes could provide an accurate estimate of the cost of the trip in terms of time, distance travelled, fuel spent, and environmental impact. As the dataset provides the source and destination of each trip, deploying an API with some rules on road accessibility can determine the truck route which would provide more detail on each, thus adding a more granular insight to the true cost of each trip and in turn how much can be saved if these trips were optimized or eliminated.