Datasheet for Open Data Toronto's "Land Ambulance Response Time Standard"*

Sebastian Rodriguez

25 March 2023

A datasheet for Open Data Toronto's Land Ambulance Response Time Standard dataset (*Land Ambulance Response Time Standard* 2022), original data available under the Open Government License – Toronto. The questions used to form this datasheet were extracted from Gebru et al. (2021). Created using the R statistical programming language (R Core Team 2020).

Table of contents

*Code and data are available at: https://github.com/seb646/toronto-paramedics-datasheet.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

    - Generally, the dataset was created to facilitate transparent government by the City of Toronto and the Toronto Paramedic Services. The dataset, hosted on Open Data Toronto, allows public inspection of the paramedic service's response times. The Toronto Paramedic Service is required by legislation to achieve a certain response time to each emergency call, determined by both the Government of Ontario and the City of Toronto. The service must submit reports of their compliance with these response times to retain its funding and certifications.

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

    - There is little information publicly available about who specifically created this dataset. It was created for Open Data Toronto using data from the Toronto Paramedic Services, although it is unclear whether someone from the service or from Open Data Toronto compiled the dataset.

3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

    - The creation of this dataset was funded by the City of Toronto.

4. *Any other comments?*

    - There is much ambiguity as to the author of this dataset. It may have been authored by Open Data Toronto or the Toronto Paramedic Services. Although the stated "publisher" is the Toronto Paramedic Services, the listed contact for the dataset is Open Data Toronto.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

    - Each row of the dataset represents a response time standard (for example, CTAS 1), its matching performance plan, and the realized response time provided by the Toronto Paramedic Services. As the dataset spans multiple years, each instance of data repeats for each year. For example, there are multiple lines of CTAS 1 data for 2013, 2014, and so on.

2. *How many instances are there in total (of each type, if appropriate)?*

    - There are 6 types of data (each representing 1 of 6 response time standards) and 8 instances of those types (each representing a year between 2013-2021, with 2020 included but missing realized data).

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

    - The dataset is a sample of instances from a larger dataset. The data used in this set are a summarized review of response times from the Toronto Paramedic Services larger performance metrics. Additionally, the breadth of this data spans only to 2013, although the service itself was created in 1975. It is unclear why this specific range was selected, and if there are records pre-dating 2013. It is also unclear whether this data is representative of historical response times, although it is representative of the service's larger performance times as it is a summary.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

    - Each instance consists of a response time standard along with its target response time in minutes, target percentage of calls served with that response time, the realized performance of the response time against the target performance, the year associated with the instance, and the date that the data was submitted.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

    - Each instance is categorized by a response time standard and a matching year to allow ease of sorting. In addition, two of the columns utilize a percentage scale that must not exceed 100% nor fall below 0%.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

    - Yes. The report labels for 2022 are included in this dataset (along with the target responses), but the realized performance is missing. This is due to delayed reporting and will most likely be available on March 31, 2023.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

    - Yes, the response time standard remains consistent across the data as these patient severity categorizations are legislated by the government. Similarly, the target response times and target performance metrics also share similarities, although there are some discrepancies that reflect changes in said legislation.

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

    - The data is split among response times of medical services to a patient and response time of communication services to a patient. This split represents two unique standards: a standard of physical care, and a standard of virtual care. Both are vital, as virtual care is required to first assess the severity of a situation and inform paramedics of the emergency.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

    - No, there are no apparent errors or sources of noise. There are redundancies, although these are purposeful and provide added meaning to the data (for example, the repeating response time standard labels and repeating performance plan targets).

10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

    - The dataset is fairly self-contained, relying only on metrics provided by the government. It does not change based on citizen-reported data. However, it will change depending on changes in legislation and regulation.

11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

    - The data does not provide any confidential information, although its summary statistics rely on information that may be confidential (such as an incident report).

12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

    - No, the dataset does not include information that would harm a viewer. However, if the realized performance of response times drops below the target performance, then it may cause rightful concern.

13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

    - No, it does not categorize calls based on sub-populations like age, gender, or region.

14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

    - No, it is not possible to identify any individuals from this dataset.

15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

    - No, the dataset does not include any sensitive information.

16. *Any other comments?*

    - No.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

    - The data associated with each instance uses a mix of information from two sources: individual responses by the Toronto Paramedic Services and legislation from both the Government of Ontario and the City of Toronto. The individual responses are reported by the Toronto Paramedic Services using internal data on the performance of each emergency response call, while the legislation is determined by provincial parliament and Toronto's municipal government.

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

    - The data recorded by the Toronto Paramedic Services were recorded using a mix of hardware and human measurement. Hardware determines when a call was first placed to an emergency dispatcher, while paramedics themselves report when they first arrive on scene.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

    - There was no sampling strategy employed. The data is an inclusive and comprehensive statistical summary of all emergency medical incidents responded to in a given year.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

    - Data collection was conducted by both emergency service dispatchers and paramedics. This collection is a required component of both collector's job and thus is not directly relevant to the dataset, although salaries of government employees are public domain.

5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*

- The data is collected continuously and there is no set timeframe. The data is divided by year, beginning at 00:00:00 on January 1st and ending at 11:59:59 on December 31 of each calendar year.

6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - It is unclear whether there was an ethical review of this data.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

    - The data was obtained via Open Data Toronto, a project of the City of Toronto.

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

    - No, as the data is anonymized and does not contain any confidential or sensitive information the individuals that form this data were not notified.

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

    - The individuals consent to data collection for internal government purposes when they place a request to use the paramedics service.

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

    - Canada's privacy laws may allow individuals to request the deletion of their data from systems, although it is unclear whether these laws apply to the government.

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - No, no analysis was provided.

12. *Any other comments?*

    - No.

Preprocessing, cleaning, and labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

    - Yes, comprehensive preprocessing and cleaning was done prior to the dataset's publication. The dataset is a summary of a larger dataset on response times for individual cases, and these individual cases were omitted from the final dataset.

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

    - The raw data was saved in a separate dataset, although it was processed separately before publication and also does not reflect all components of the original dataset.

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

    - No, it is unclear what software was used to preprocess, clean, and label the data.

4. *Any other comments?*

    - No.

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - There is no feasible way of determining how the data was previously or is currently used. The Open Data Toronto allows any person to utilize the dataset. One project known to have used this dataset is an analysis of the Toronto Paramedic Services' response times, available here: https://github.com/seb646/toronto-paramedic-responses.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
    - No, there is no public log of any papers or systems that currently use the dataset.
3. *What (other) tasks could the dataset be used for?*
    - The data set can be used to publicly inspect the Toronto Paramedic Services' response time standards and their realized performance of those standards.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
    - No.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
    - No. The purpose of Open Data Toronto is to allow the public to use government data as they see fit.
6. *Any other comments?*
    - No.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
    - The dataset is freely distributed on the Open Data Toronto website, found here: https://open.toronto.ca/dataset/land-ambulance-response-time-standard.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
    - The dataset is distributed on the Open Data Toronto website.
3. *When will the dataset be distributed?*
    - The dataset has no set schedule for distribution or updates, although previous data suggests updates on March 31st of each calendar year.
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
    - The dataset is provided and licenced under the Open Government Licence - Toronto, found here: https://open.toronto.ca/open-data-license.
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
    - No, the data is government-owned and government-published so no third-parties have a legal claim to the data.
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
    - The dataset is governed by the aforementioned Open Government License provided by the City of Toronto, and may be further restricted by municipal, provincial, or federal legislation.
7. *Any other comments?*
    - No.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
    - The Toronto Paramedic Services maintain and update the data, while Open Data Toronto supports and hosts the dataset.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - The manager of the dataset may be contacted through Open Data Toronto via opendats@toronto.ca.
3. *Is there an erratum? If so, please provide a link or other access point.*
    - No.
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
    - Yes, the dataset will be updated *at least* once per calendar year on the 31st of March with new data from the Toronto Paramedic Services. The updated dataset will appear on Open Data Toronto's website, although there is no formal process to notify consumers of updates.
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
    - No.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
    - Yes, sort of. The data from previous datasets will be reflected in new versions. Although previous versions may not be retained for public consumption.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
    - With open data, there are always mechanisms for others to extend, augment, or build upon the dataset. However, there is no formal mechanism to bring these

changes to the Open Data Toronto platform or the official dataset. Users may take the data set and augment it on their own, provided they comply with the Open Government Toronto license.

8. *Any other comments?*

    - No.

References

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Land Ambulance Response Time Standard. 2022. Open Data Toronto. https://open.toronto.ca/dataset/land-ambulance-response-time-standard.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.