UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

datamanagementgroup

2016-2018

Data Management Group

Activity Report

Prepared by:

Data Management Group
Department of Civil and Mineral Engineering
University of Toronto

March 31, 2019

INTRODUCTION

The Data Management Group (DMG) was established in 1988 on the basis of a proposal from the University of Toronto's Joint Program in Transportation for an autonomous research group with the following objectives:

(a) to establish a common, centrally accessible database containing information on transportation activities, zone systems, transportation networks and land use activity,
(b) to provide a transportation data retrieval service to the participating agencies,
(c) to monitor the adequacy of available data and propose approaches for adding to or updating the data as mutually agreed upon by the agencies,
(d) to promote greater interaction between university researchers and practitioners in the field of urban transportation planning,
(e) to promote the communication of transportation information and data obtained or administered by the Data Management Group to interested agencies and to the public, and
(f) to further the improvement of transportation demand analysis, research, and forecasting in the Greater Toronto Area.

Although the administration of the group has changed and DMG is now a part of the University of Toronto Transportation Research Institute (UTTRI) under the Department of Civil and Mineral Engineering at the University of Toronto, the DMG continues to be guided by these objectives into this its 30th year of continuous operation.

Program approval and funding of the DMG is the collective responsibility of members of the Transportation Information Steering Committee (TISC) with the following membership:

City of Hamilton
City of Toronto
Metrolinx
Ministry of Transportation, Ontario
Regional Municipality of Durham
Regional Municipality of Halton
Regional Municipality of Peel
Regional Municipality of York
Toronto Transit Commission

Each participating agency appoints a member of their technical staff to the Transportation Research and Data Management Group (TRADMAG), which is a

standing committee of TISC, and is responsible for coordinating the needs of the funding agencies and the activities of the research project.

This report provides a brief profile of the staff employed and a description of the activities undertaken by the DMG during the calendar years 2016 to 2018, inclusive.

STAFF AND LOCATION

Director

Eric J. Miller, Professor, Department of Civil & Mineral Engineering, University of Toronto

Technical Staff

Reuben Briggs, B.A.Sc. (Civil Engineering), M.A.Sc. (Civil Engineering) University of Toronto, P.Eng.

Jason Chan, B.A.Sc. (Electrical & Computer Engineering) University of Toronto (until August 2016)

Susanna Choy, B.A.Sc. (Industrial Engineering), M.Eng. (Civil Engineering) University of Toronto, P.Eng.

Asmus Georgi, Vor-Diplom (Geography), Diplom (Geography, minors in Chemistry and Transportation Planning & Engineering), Freie Universität & Technische Universität, Germany (since August 2017)

Summer Student

Chantal Wong, 2nd year undergraduate, Department of Civil engineering, University of Toronto (2017)

Web Site

http://www.dmg.utoronto.ca

Office Location

Department of Civil and Mineral Engineering
University of Toronto
Galbraith Building, Suite 305
35 St. George Street
Toronto, Ontario M5S 1A4
Telephone: (416)978-3913
Email: info@dmg.utoronto.ca

TRANSPORTATION TOMORROW SURVEY

Under the guidance of TISC, a series of urban travel surveys, Transportation Tomorrow Survey (TTS), have been conducted every five years since 1986. The last TTS was conducted in 2016. Previous surveys since 1991 had been managed by the Data Management Group but due to the absence of a principal investigator from the University of Toronto in 2016, the DMG participated in the 2016 survey as an advisor to TISC and provided consulting and support services for the chosen vendor.

DMG participated in the survey vendor review and selection process in early 2016 and once the vendor was chosen became part of the Technical Advisory Committee (TAC) assigned by TISC to monitor the vendor's progress.

DMG played the role of a technical resource to the vendor, playing various roles throughout the stages of the survey. DMG provided training manuals and maps to the vendor and provided technical and computer resources for the usage of the Direct Data Entry (DDE) survey software. The DDE which had been developed by DMG had been used in the last two previous surveys and was used in 2016 to conduct the telephone portion of the survey. DMG was responsible for its installation and running at the survey site during the data collection portion of the survey. As well DMG offered advice and reviewed the CallWeb web survey tool which was developed by the vendor to handle the online portion of the survey.

DMG also observed and advised the vendor's management team at the survey call-centre in Toronto, doing quality checks on random survey response samples, reviewing survey timelines and also providing additional resources to the special team to review transit data collected in the surveys.

During the validation and report creation stage, DMG extracted the data from DDE for provision to the vendor to create the final data set, provided the vendor with previous survey summary data for report creation and reviewed all vendor created reports.

The TTS data was released to the public in August 2017.

INFORMATION PROCESSING

The term 'information processing' is used in this instance to describe a set of activities supporting the management, storage and distribution of urban travel information. The principal components of this information are the results of the Transportation Tomorrow Surveys and a collection of all Cordon Count Surveys.

Transportation Tomorrow Surveys and iDRS

The DMG administers the data files on urban travel contained in the 1986, 1991, 1996, 2001, 2006, 2011 and 2016 Transportation Tomorrow Surveys (TTS) in the form of a set of relational databases with various methods of access. Direct access to the original files is restricted to DMG staff to ensure that information on a particular household cannot be identified. Currently, data files available to iDRS users contain the following information:

TTS Number of records

Year	Households	Persons	All Trips	Transit Detail
1986	61,453	171,086	370,248	56,615
1991	24,507	72,496	157,349	14,896
1996	115,193	312,781	657,971	702,95
2001	136,379	374,182	817,744	85,095
2006	149,631	401,653	864,348	87,244
2011	159,157	410,404	858,848	86,703
2016	162,708	395,885	798,093	91,437

The increasing size of the databases reflects not only growth in the area but also changes in the size of the area surveyed.

The first text-based data retrieval system (drs) was developed in the mid-1990s by the staff of DMG as the original method for external users to gain access to the data files and complete the data extraction themselves. This retrieval system was very effective when a modem was used as the principal method of remote access to the DMG's computer system. Over the years, as the demand for travel data grew and the Internet became the preferred method of remote access, a data retrieval system specifically designed for Internet access was developed (iDRS). All the features of drs were incorporated into the browser-based iDRS and the drs process was phased out.

The initial release of iDRS was restricted to use by the funding agencies. In 2002, access to iDRS was made available to any individual that requested access. The individual was required to sign an agreement form and system security was

maintained by giving each user a unique login and password. This procedure has the added benefit that agencies outside the Greater Toronto and Hamilton Area that participated in the Transportation Tomorrow Surveys could access their data without the need to set up their own database system.

In order to meet the changing needs of the data, the DMG developed a new version of iDRS. The software was rewritten to incorporate a new database management system; PostgreSQL, which was used as part of the TTS survey software. Several new functions such as additional filter criteria and the ability to store the query for future use have been added and the speed of extraction has increased dramatically.

A new authorization process has also been implemented together with the new software. The process has been automated so that any individual can request access by entering the name, affiliation, email address and contact number online. Once the information is submitted, it will be approved by the DMG staff and an email with the credentials will be sent to the new user.

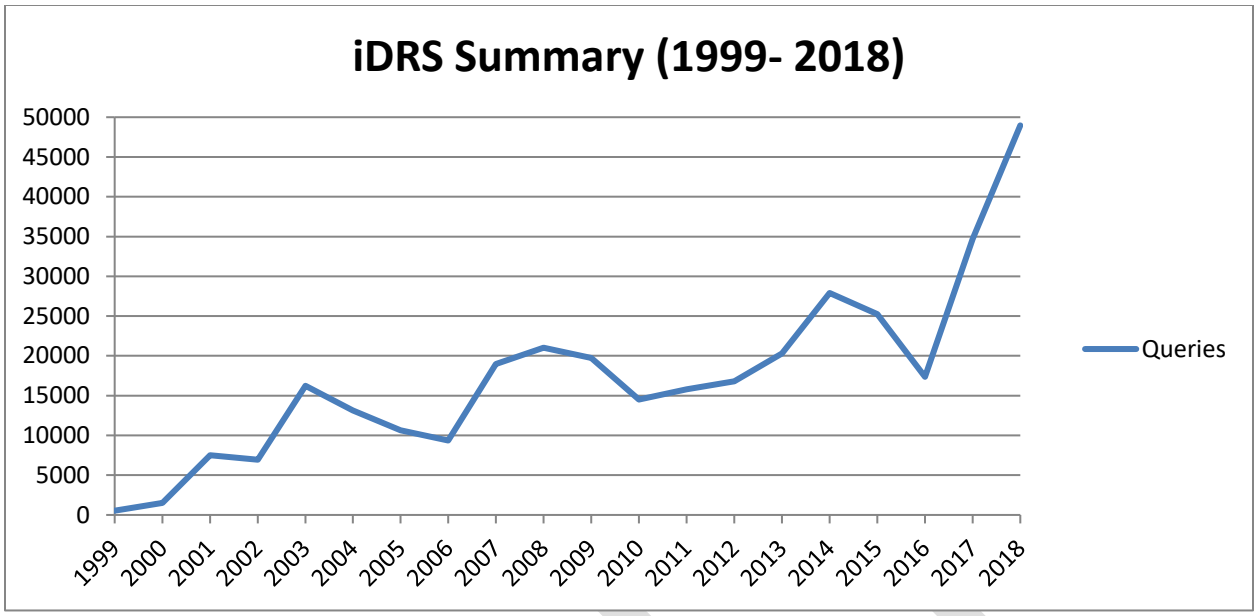
The official version of the new iDRS was released in 2016 but the DMG continues to fine-tune the new iDRS system based on the feedback of the users.

A History of iDRS Data Requests

When the browser-based data extraction procedure (iDRS) was first released in 1999, the users were registered users of the DMG's computer system. The initial use of iDRS was encouraging. Subsequently in late 2000, the Steering Committee gave approval for use by anyone registering and providing information on their affiliation. The growth in use of TTS data is reflected in the growth of the use of iDRS for data extraction. The following chart shows the growth since iDRS was first introduced in 1999. The 'Number of Queries' reflects the number of times an output was generated during a session. iDRS usage peaked at the year after a new TTS dataset was released in 2002, 2007, 2014[1] and 2018[2]

[1] The interview phase of the 2011 TTS completed in December 2012 and the final data set was released in November 2013.

[2] The interview phase of the 2016 TTS was completed in December 2016 and the final data was released in October 2017.



The chart shows significant increases in iDRS usage in the past few years.

Summary of iDRS Data Requests in 2016-2018

In 2016, there were 17,377 iDRS queries, in 2017 34,681 and in 2018, 48,970. The steep increase coincides with the anticipation of the release of the 2016 data. The following table shows the monthly summaries for all three years.

	Year		
Month	**2016**	**2017**	**2018**
January	235	2194	3079
February	426	2179	4126
March	369	4724	5479
April	177	2115	4080
May	1745	4517	7466
June	1928	2259	4944
July	2049	2493	5405
August	1736	2391	3139
September	2361	3689	3196
October	2299	3172	2759
November	2382	2889	3148
December	1670	2059	2149

Affiliations of iDRS Users in 2016-2018

The following is a list of 167 different agencies and groups that extracted the data through iDRS in 2016-2018.

407 ETR	AECOM
Alberta Transportation	Amec Foster Wheeler
Arup	Associated Engineering
Astral	AVARTZ
BA Group	Brampton Cycling
Bramptonist	Canadian Urban Institute
Carleton Univeristy	CC Tatham
CFC	CIMA
City of Barrie	City of Brampton
City of Brantford	City of Burlington
City of Edmonton	City of Guelph
City of Hamilton	City of Markham
City of Mississauga	City of Peterborough
City of Toronto	City of Vaughan
City of Waterloo	ClickInsight
Cole Engineering	Community Development Halton
Concordia University	County of Simcoe
CPCS	Cycle Toronto
Deloitte	DevTrans Engineering
Dillon Consulting	DMG
Durham Region Transit	Entertainment District
Environics Analytics	EXP Services
Frends of the Greenbelt Foundation	GDA
General Motors	GHD
Gladki Planning Associates	Golder Associates
Government of Ontario	Greater Toronto Airports Authority
Hafencity University	Halton Region
Hatch	HDR
Heart of Newmarket	Hemson
IBI	Independent consultant
INRO	Jakarta
Jarrett Walker and Associates	JBK
Laval University	LEA
MaRS Discovery District	McGill University
McKinsey and Company	McMaster University

Memorial University of Newfoundland	Metrolinx
MHBC Planning	MIC
Ministry of Municipal Affairs	Mohawk College
MTO	Newcastle University
NexTrans Engineering	Niagara Connects
NTNU - Norway	Open Government office
OPPI	Ortech
Paradigm Transportation Solutions Ltd	Parsons
Parsons	Parsons Inc
Pembina Institute	Peoplecount
Personal	Polytechinique Montreal
Porche Canada	Poulos and Chung Ltd
Private residents	PTV Group
Public	Public policy Consultatn
PWC	Queens University
QUT University	R J Burnside
R. C. Lau Inc	Read Voorhees and Associates
Realtor	Region of Durham
Region of Niagara	Region of Peel
Region of Waterloo	Region of York
Ryerson University	SDG World
Seneca	Sidewalk Labs
Social Planning Toronto	Spectrum Traffic Data Inc
SPRC Hamilton	Spring Farm Ratepayers Association
SSG	St. Michaels - UHS
Stantec	Steer Davies Gleave
SUDA	Sustainable Severn Sound
TCDSB	Technical University of Munich
The Municipal Infrastructure Group	The Neptis Foundation
Thompson Ho Transportation Inc	TIS
T-Kartor Sweden AB	Toronto Atmospheric Fund
Toronto Public Health	Toronto Star
Torrie Smith, TAF	Town of Caledon
Town of Milton	Town of Newmarket
Town of Oakville	Town of Richmond Hill
TrafficPlus Engineering Ltd.	Tranplan Associates
Trans-Plan Transportation Inc.	Trent University
Tribhuvan University	Tri-Cities Transport Action Group
TTC	UEM Consulting
University of British Columbia	University of Guelph
University of Manitoba	University of Toronto

University of Waterloo	University of Windsor
UOIT	UPM
Urban Strategies Inc	UW - Madison
Valcoustics Canada Ltd	VIA Rail
Vividata	Volkswagen Group
Waterford Institute of Technology	Wellesley Institute
WhatIf Technologies	WSP
York University	

TTS Special Data Requests

The interactive procedures available with iDRS satisfy the majority of data needs. However, some data needs are too complex and require the intervention of an experienced analyst to formulate a custom query from the database. In addition, the DMG staff can often help define the most relevant data for the problem at hand. One of the typical data requests involves assigning the x-y coordinates of households or trip ends to a specified spatial aggregation different from the predefined traffic zone system as these coordinates are not available to any non-DMG staff in order to protect the privacy of the survey respondents. There is also a growing interest in using disaggregate data for travel forecasting among academic researchers. Here is a list of agencies that have made a special data request from 2016 to 2018:

Cancer Care Ontario
Centre for Urban Health Solutions/St. Michael's Hospital
City of Toronto
City of Markham
City of Peterborough
Metrolinx
Ministry of Transportation Ontario
Regional Municipality of Durham
Regional Municipality of Halton
Regional Municipality of Waterloo
Regional Municipality of York
University of Montreal
University of Toronto

Cordon Counts and CCDRS

The City of Toronto (then the Regional Municipality of Metropolitan Toronto) began collecting detailed information on the type and volume of traffic crossing selected points on the road system as early as 1975. The counting locations were selected such that screen lines or cordon lines could be defined and the counting program has continued on a regular basis since that time, usually twice in a five year cycle.

Subsequently, other Regions began similar programs. Given the number of Regions with a similar program, they began coordinating their count programs and defining a common set of data standards. In 1998, the DMG collected these traffic counts in a common database structure and developed a Cordon Count Data Retrieval System (CCDRS). CCDRS is now widely used by a variety of public and private agencies.

In addition, participating agencies now use CCDRS as a tool in verifying their cordon count results. The approved procedure is to place new cordon count data directly into the CCDRS database with a disclaimer notice to all users that the new information is preliminary. The last Cordon Count was undertaken by the participating regions in 2016.

The 2016 cordon counts were partially released at the beginning of 2017 with data from the City of Hamilton. The remaining data from the other Cities and Regions of Durham and York was released by the end of summer 2017.

The DMG released a beta version of a new CCDRS in 2016. The final product is still a work in progress and is to include a graphic interface so the users can identify and select screen lines and count stations from a map. Statistics for usage of the new CCDRS are unavailable at this time.

Open Data

In March 2011, the first generation of Open Data Portal was launched by the Government of Canada. "Open Data is defined as structured data that is machine-readable, freely shared, used and built on without restrictions." The data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. It must also be available in a convenient and modifiable form and must be provided under terms that permit re-use and redistribution including the intermixing with other datasets. Everyone must be able to use, re-use and redistribute. There should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g., only in education), are not allowed.

With the approval from TISC, the DMG opened up the TTS and cordon count data to the public in 2014. Unlike the online data retrieval systems, users can download the pre-generated text files from the DMG web site without registration.

The Open Data Portal at the DMG was updated to include the 2016 TTS data and the 2016 Cordon Count data bringing the total number of files available to 129 TTS data files and 222 Cordon Count data files. Each TTS data file contains household, person, and trip information for a specific survey year, geographic area and spatial aggregation

for different time periods. Each cordon count data file contains different types of vehicle and person counts for a specific year, geographic area and time period. The files are in comma-delimited text format and readily import into Excel or other spreadsheet software. The DMG continues to update the Open Data Portal to include new TTS and cordon count data files.

Disaggregate Data Access

More and more transportation related projects are requiring access to disaggregate data for modelling work. With approval from TISC in 2017, DMG has begun testing a new data access protocol for the access of disaggregate data by researchers and consultants working on projects on behalf of DMG's funding partners which will allow DMG to maintain security over the disaggregate data.

Potential users are required to fill out a request form which outlines the specific data required, project and its sponsor, and the time frame for which access is required. If the request is approved, an account is set up and the users are given login instructions and a unique login and password to access their required datasets on a Virtual Machine. Datasets provided do not include any x-y coordinates and all locations are coded to a higher-level geographic area i.e.: traffic zones, Census dissemination areas, planning districts etc.

The server is setup so the user can work on the datasets in the account via Remote Desktop Protocol (RDP). However, there is no internet access allowed and data cannot be directly copied off the server. Only aggregate data is allowed off the system and any data files required to be downloaded must be reviewed and approved by DMG personnel who will then allow download of the data from an FTP location. DMG has committed to review and forward requested data within one business day.

Below is a list of the users of this new data access in 2017 and 2018:

Company	**Agency working on behalf of**	**Data required**	**Project**	**Start Date**	**End Date**
Arup	Metrolinx	TTS 2016	Transit Oriented Development Ridership Forecasting study	April 2018	August 2018
Arup	Metrolinx	TTS2016	Brampton Queen St - York Regional Hwy 7 BRT Planning Study	July 2018	March 2019

Steer	Metrolinx	TTS 2016	Fare Integration Ridership Response Model	March 2018	December 2018
WSP	York Region	TTS 2016	York Regional Model	January 2017	December 2019
WSP	Halton Region	TTS 2016	Halton Regional Model	October 2018	April 2021
Prof. Steven Farber	University of Toronto	TTS 2011, 2016	Various research projects	January 2017	September 2022

COMPUTER RESOURCES AND TECHNICAL SUPPORT

The DMG computer system is comprised of several servers including the DMG main server, the data retrieval server, an EMME licence server, a backup server and a file server where the data files of the funding agencies are stored. All of these servers are located behind the DMG firewall for security reasons.

The DMG main server is a Dell R620 running Windows Server 2012 R2 with Hyper-V application. This application allows the DMG server to run multiple virtual machines, i.e., operating systems emulated within another operating systems by imitating dedicated hardware.

There are five virtual machines running on the DMG server. The first one is the mail server using Microsoft Exchange Server 2013 to handle daily DMG emails. The second one is the DMG website created in WordPress. Linux, Apache, MySQL, PHP and Lamp stack are running on this machine. The third virtual machine is the new iDRS developed using the MEAN stack, a free and open-source JavaScript software stack for building web applications running in Debian. The data are hosted in PostgreSQL which is an open-source relationship database management system. The use of open-source software allows free access, modification and sharing of the software. In 2016, a switchover was made from the old iDRS legacy system to this new system which was previously in beta release.

The data retrieval server hosts the TTS and Cordon Count. All years of the TTS and Cordon Count Data are compiled and stored in a consistent way, and the TTS and Cordon Count data are frequently used together in regional planning projects or traffic impact studies. A summary of the use and the users of these services is contained earlier in this annual report.

The fourth and fifth virtual machines host the disaggregate data access. Consultants and Researchers can access disaggregate data for their projects and DMG allows them to export aggregated outputs. These two machines are:

- A Dell R430 server, running Windows Server 2012 R2 Datacenter, and Hyper-V with an Intel Xeon CPU E5-2650L v3 (x2) processor and 160GB of Memory and 2x1TB of disk space.
- A Dell R620 server, Microsoft Windows Server 2012 R2 Datacenter with an Intel Xeon CPU E5-1690 (x2) processor, 64 GB o Memory and 1.5TB of disk space

The EMME software, which is used by the funding agencies for modelling, is run locally on a personal computer with all related files stored locally. This results in a much faster operation. Each local machine requires access to an authentication key to operate the software. Rather than each agency purchasing a licence from INRO to use

the software locally, DMG negotiated a concurrent licence for authentication of several machines operating at the same time. The DMG then dedicated a server to provide remote authentication to the participating partners. The DMG is in the process of migrating the current suite of supporting hardware and software to more recent and cost-effective alternatives.

The EMME software completely replaces the previously used UNIX based EMME2 software and the old EMME2 server was decommissioned in 2018 as INRO ceased support of that software and all EMME users migrated to the PC-based EMME version.

The DMG continues to maintain and improve the computer system to meet its funding agencies' changing needs.