



EXPLORING DATA WRANGLING PROTOCOLS IN MAPPING APPLICATIONS

Greice C. Mariano, Sara Diamond

© OCAD University :: Visual Analytics Lab :: 205 Richmond Street West Suite 720, Toronto, Ontario.

1. Urban Informatics for transportation operations, planning & decision-making



3. Visualization & Decisionsupport 2. Urban Mobility & Integrated urban systems design



O C A D





O C A D U





Research Approach

Reference Model for Visualization

O C A D

U



Image adapted from: Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. Readings in information visualization: using vision to think.

Geographic Information System (GIS)

- Is a computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface.
- Can help the individuals and organizations better understand spatial patterns and relationships
- Requires a georeferenced data:
 - Can use any information that includes location, which can be expressed in many different ways, such as latitude and longitude, address, or zip code.



Views for Urban Informatics - Compara



O C

A D

U

C engineer C citizen O designer architect 🕤 urban planner Analyst O transportation engineer O researcher O civil engineer mobile app user O concept development C technician O urban designer O planners for consultation C computer scientist real estate developer O surveyor O consultant O product designer O 3D rendering O CAM O transportation professional O design consultant O elected official O community user O cartographer O planners with specialized training O concept developer O risk management O defence O telecommunications O GIS O location-based services O data visualization O visualization O city planner O AEC

Views for Urban Informatics - Compara

O C A D

Research Overview	Urban Design and Planning Software 🔘	Autocad and Autocad Civil 3D - Autodesk Revit - BIM, Autodesk 3DS Max, Autodesk Sketchup O Sketchup Viewer O Vectorworks O ESRI / ArcGIS O Betaxille O UrbanSim, Urban Carvas (Autodesk) O UNA Toolbox- MIT StreetFactory O Fusion 360 (Autodesk) O	 ⊖ engineer ⊖ citizen ⊖ designer
	Design Software 🔾	Solid WORKS / Dassault Systemes O Blender	architect
Urban Design - Built Environment, neighbourh	ood planning	Urban Design and Planning Software O Design Software O	Autocad and Autocad Civil 3D - Autodesk Revit - BIM, Autodesk SDS Max, Autodesk Sketchup my.Sketchup Sketchup Viewer Vectorworks Microstation ESRI / ArcGIS Betaville UrbanSim, Urban Canvas (Autodesk) DINA Toolbox- MIT StreetFactory Fusion 360 (Autodesk) Solid WORKS / Dassault Systemes Blender CityGML Stamen map generation openstreet map Nodex Blender GIS Unfolding Maps MapBox ViziCities Cart DB
		StreetWix O Textizen	
Infrastructure Management, Sustainability, and Resilient cities 🔘	Infrastructure Management	Cit'Ease SmartCity O	 Gis Iocation-based services data visualization
Simulation / Interaction / Games	Simulation / Interaction / Games 🔘	UrbanSim O Ecopolicy Game simulation	
Big Data & Analysis 🔿	Big Data & Analysis 🔵	Watson Analytics 🔘	O AEC







O C A D













O C A D











O C A D

U





13

GIS Representation

- Vector:
 - Data is represented using points, lines and polygons
- Raster:

O C A D

U

- Data is represented as a surface modeled by a matrix of values (pixels)
- Useful for continuous data, such as satellite imagery, aerial photographs, pollution, population, etc.



Image source: http://3.bp.blogspot.com/-A140pKBwSXU/VHvp73TcocI/AAAAAAAAAAAK8/xoP1KQI5L-Y/s1600/Raster%2Band%2BVector%2BData.jpg

GIS File Formats



Image source: https://gisgeography.com/qgis-arcgis-differences/

O C A D

GIS File Formats



Image source: https://gisgeography.com/qgis-arcgis-differences/

O C A D











O C A D

U





17

Research Approach

Reference Model for Visualization

O C A D

U



Image adapted from: Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. Readings in information visualization: using vision to think.

Data Wrangling

"It is the process of **converting** and **mapping data** from its **raw** form to **another format** with the purpose of **making it more valuable** and **appropriate** for advanced tasks in Data Analytics and Machine Learning."



Use Case

iCity Working In Progress

- Transportation Tomorrow Survey (TTS)
 - Data 1: Origin-Destination Matrices
 - File format: set of CSVs files
 - Each file refers to a different variables, such as cost, time wait, volume of people...
 - Data 2: TTC zones Shapefile
- Visualization using Kepler.gl and Deck.gl
 - Steps:
 - 1. Transformation of shapefile to GeoJson
 - 2. Data transformation from OD-Matrices to Data Tables
 - 3. Integration between the different variables in the Data Tables
 - 4. Mapping the new data tables with the TTC zones.

Shapefile

- ESRI standard vector file format
- Enconding features as: point, lines and polygons
- Stored as a set of related files and contains one feature class.

GeoJson

- Open standard vector file format based on JavaScript Object Notation (JSON)
- Encoding features as: point, lineString, polygon, MultiPoint, MultiLineString, MultiPolygon, and GeometryCollection.
- Features contains: geometry object and additional properties, and a FeatureCollection contains a list of Features.

O C A D

Shapefile → GeoJson



O C A D U Images source: http://www.gisresources.com/understanding-shapefile-shp-file-format/ (shp) and http://chittagongit.com/icon/json-file-icon-12.html (json)

Shapefile x Geojson





OD-Matrices to Data Tables and integrated with shapefile

OD-Matrices - Directory

AM		📄 Auto	Ø 🕨	bpenalty.csv	0
EV	Ø 🕨 📘	Transit	🥝 🕨	ccost.csv	0
MD	🛛 🕨			distance.csv	0
PM	🥥 🕨			ptt.csv	0
				tfare.csv	0
				tivtt.csv	
				transit_time.csv	0
				true_transit_time.csv	0
				twait.csv	0
				twalk.csv	0



OD-Matrices to Data Tables and integrated with shapefile

OD-Matrices - Directory

AM	S	📃 Auto	Ø 🕨	bpenalty.csv	0
EV	Ø 🕨	🔲 Transit	🥝 🕨	ccost.csv	0
MD	Ø 🕨			distance.csv	0
PM	🥥 Þ			ptt.csv	0
				tfare.csv	0
				tivtt.csv	
				transit_time.csv	0
				true_transit_time.csv	0
				twait.csv	0
				twalk.csv	0



OD-Matrices to Data Tables and integrated with shapefile

trans	sit_time.c	sv				
Zones O\D	1	2	3	4	5	6
1	0	29.7761	21.25672	24.78801	29.80301	38.22379
2	30.92986	0	22.35746	30.90995	35.72341	36.07895
3	20.17063	22.80784	0	20.66859	25.51105	32.99121
4	23.45448	27.36932	18.84467	0	21.38223	22.12184
5	28.57211	34.18865	24.28002	20.88825	0	22.06426

O C A D

V	4_zones.g	eojson					
OBJECTID int	gta06_net0	ORIG_FID	Shape_Leng	Shape_Area	FID_	NUM	GTA06
1	1	1	3240.40115051	523608.711076	2151	1	1
2	2	2	10264.6422225	1179813.05658	1826	2	2
3	3	3	2149.07481928	220460.93084	2153	3	3
4	4	4	2439.97969026	253035.243478	1961	4	4
5	5	5	3050.44455033	501072.324501	1960	5	5
6	6	6	5489.01453565	1054622.94564	1939	6	6







iCity Working In Progress

- Visualizations in Kepler.gl and Deck.gl
- New design and prototypes for:
 - Complete Streets dashboard
 - Transit Management dashboard
 - iCity ORF Integrated Website

Acknowledgements

- Ontario Research Funding (ORF)
- NSERC
- Esri Canada.



References

- <u>https://journals.sagepub.com/doi/full/10.1177/1473871611</u>
 <u>415994</u>
- https://www.digitalvidya.com/blog/what-is-data-wrangling/





THANK YOU!

Find out more about research at OCAD U at: <u>http://www.ocadu.ca/research</u>



© OCAD University :: Visual Analytics Lab :: 205 Richmond Street West Suite 720, Toronto, Ontario.