

Master Thesis



Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of computer graphics and interaction

Urban Data Visualization

Semestral Project Report

Vojtěch Tomas

Supervisor: Ing. David Sedláček, Ph.D.
January 2021

Acknowledgements

Děkuji ČVUT, že mi je tak dobrou *alma mater*.

Declaration

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškerou použitou literaturu.

V Praze, 14. January 2021

Abstract

TODO

Keywords: word, key

Supervisor: Ing. David Sedláček, Ph.D.

Abstrakt

TODO

Klíčová slova: slovo, klíč

Překlad názvu: Vizualizace
urbanistických dat — Zpráva k
semestrálnímu projektu

Contents

1 Introduction	1
1.1 From Data to Visualization	1
1.2 The Goals	2
2 Urban Data	5
2.1 Urban Informatics	5
2.2 Sources of Urban Data	5
2.3 Urban Data Domains	6
2.4 File Formats	6
2.5 Models Based on Urban Data . . .	8
3 Existing solutions	11
3.1 Visualization Tools with Physical Components	11
3.2 Purely Virtual Visualization Tools	12
A Bibliography	13

Figures

1.1 Basic city planning loop	1
2.1 Sociotechnical system model [1] . .	8
3.1 Tactile Matrix framework	11

Tables

2.1 Urban data classes availability on open data portals	7
---	---

Chapter 1

Introduction

In recent years, cities have become a huge source of various data. The urban data usually includes street layouts, public transport information, layouts of the power network, noise maps, socio-economic maps, etc. The first natural question is, why should we care about the urban data? Why do we bother coming up with creative ways to present the data publicly?

All plans and future changes are made and approved based on the models derived from the urban data. As the city develops, people adapt, and newer data is generated. Newer data can help building newer, better models. The city planning process is a loop with city planners on one side and stakeholders together with general public on the other, see illustration 1.1.

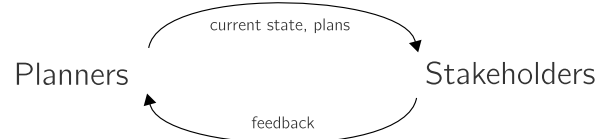


Figure 1.1: Basic city planning loop

Looking at this loop from the stakeholders' perspective, it is in their best interest to influence the decision-making process in their favour. Similarly, local authorities should strive for more interaction between the city administration and the public. This involves making the urban data publicly available and accessible.

1.1 From Data to Visualization

As of today, the urban data is usually freely accessible in a format that is obscure for the general public. Naturally, to make the data more accessible, it is necessary to visualize it. Local authorities usually make this effort with either one of the following goals:

- to inform the public about the current state,
- or present the effects of a future change in the area.

Communicating the current state usually requires analysis and simplification of the raw data. There is a wide range of tools for that; however, these tools are usually proprietary, and the outputs are not that engaging. The visualization is rarely customizable; proprietary formats are used throughout the entire pipeline, and it is impossible to combine various types of data.

Communicating the future changes is even more difficult since it is required to inform the public about the current state first. In this context, the interactivity of the presentation can be the key to understanding.

We are presented with several technical challenges. The input data is often in several incompatible formats, and each describes a completely different urban domain. In other words, the datasets are often complementary, and the visualization system has to support their combination. For example, some data use different coordinate systems, or the datasets differ in their dimensionality.

If we want to create a *usefull* visualization, the aforementioned technical obstacles should remain hidden from both the visualization creator and also the audience. In the ideal case, the visualization creator is given absolute freedom. The visualization system does not pose any hard restrictions on what is possible; the creator's imagination is the only limit. This utopistic idea is often contradicted by the requirement of a simple and usable interface. It is appropriate to seek a balance between unlimited possibilities and usability.

Taking into account the viewer's perspective, the goal is to make the visualization immersive enough to captivate, educate, and entertain. The physical setup and software capabilities should not diminish the experience and should avoid posing any virtual barriers between the medium and the audience.

Usually, the visualization is presented in a public space, either in museums or centers specifically equipped for this kind of presentation. While some of these places own the necessary equipment, the production pipeline today is often complicated, and an ad-hoc solution is used. In other cases, the visualization is presented using proprietary software, which is not primarily designed for the general public.

1.2 The Goals

The goal is to develop a modular system, which will streamline the urban data visualization process. The design should focus on the following four factors:

- the actual communicated information,
- the nature of the available urban data,
- the creator's artistic vision,
- and the equipment used for the presentation.

The tricky part is that probably only one or two of those factors are known in advance — the urban data and the used equipment can be somewhat standardized. The remaining two factors are usually project-specific; therefore, it is necessary to explore the combination of various data types and enable the freedom of artistic expression. The goal is to build a system for building other systems because of the desired modularity and adjustability — there is an inherent metaprogramming element.

Chapter 2

Urban Data

This chapter aims to describe the nature of urban data — the sources, classification, and available formats. To some extent the term *urban data* is somewhat similar to other buzzwords, such as *big data*. As there is no official definition of the term *urban data*, it is often explained by giving a list of examples, such as transport data or socio-economic maps. The urban data is a subject of study in several fields, one of which is the Urban Informatics.

2.1 Urban Informatics

In 2011, Forth et al. [2] described Urban Informatics as a separate field of study. It focuses on three key aspects — place, technology, and people in the context of urban environments. The urban environment is described as a *"complex techno-social network; the city only meaningfully exists when a sustained stream of people occupies it"* [2]. The authors outline and study four dominant trends: the emergence of ubiquitous computing, the accessibility of real-time information, informed sustainability, and planning based on citizen participation.

One of the goals of Urban Informatics is to help developing communication channels between local authorities and citizens. Moreover, thanks to the gathered information, the citizens can, directly and indirectly, influence the development of the urban area. The communication channels, as understood by Urban Informatics, are omnidirectional.

2.2 Sources of Urban Data

This section contains description of three main urban data sources. This classification is based on the summary presented in [3]. The urban data sources include mainly governments and local administrative institutions. Additionally, the data is gathered and received through devices that are part of the ubicomp device system. The third source of urban data consists of online services, such as social networks.

Open Data. The most prominent source of data are governments and local institutions. The data is commonly released in several open formats, which are further described in the section 2.4. This data often includes maps, building layouts, terrain models, technical metadata, public transport data, etc. The formats range from generic CSV or SQLite-based to domain-specific DWG, CityGML, etc. The data presented earlier are mainly static; some cities also offer real-time information about weather or public transport vehicle locations.

Ubicomp. Mark Weiser first introduced Ubiquitous Computing in 1991 [4]. The concept of ubicomp relies on small embedded computers, which communicate together and allow for seamless interaction between users and technology. According to [5], the ubiquitous information's impact on the city level became the focus of study in a separate field called Urban Computing. In the spirit of ubiquitous computing, there is an emerging field of Participatory Sensing [6], which enables gathering data from devices of individual users. The design of the participation model focuses on data credibility, security, and privacy protection. One of the applications of Participatory Sensing is the use of the gathered data for urban planning.

Online services. Personal data can also be acquired from social networks or mobile phone carriers. Online platforms usually support geotagging, adding geospatial metadata to the content posted online. The content can be easily queried based on the selected location. Alternatively, the citizens' location can be tracked via their mobile phone directly by carriers using the individual base transceiver stations (BTS). From the perspective of Urban Informatics, the aforementioned data sources enable creating richer urban area models.

2.3 Urban Data Domains

There is no official classification of the urban data. However, open data portals usually use custom classification schemes, which are either domain or format-based. Based on the available open data catalogues of the cities in 5 Cities Connect¹ group [7], a list of domain classes has been assembled, see table 2.1. The listed domains are either of significant importance or have appeared across a majority of the reviewed catalogues.

2.4 File Formats

This section describes the file formats used to store various urban data. Some of these file formats are proprietary; however, from the perspective of application development, open formats are more attractive.

¹Prague has been recently invited into the 5CC initiative

Domain	Hamburg	Vienna	Rotterdam	Helsinki	Singapur	Prague
Buildings	x	x	x	x		x
Population and Work	x	x		x	x	x
Technical infrastructure	x	x		x	x	x
Transport	x	x	x	x	x	x
Geology	x		x			x
Noise maps		x		x		x
Environment and Nature	x	x	x	x	x	x
Ortographic maps	x	x	x	x	x	x
Public services	x	x	x	x	x	x
Terrain		x	x	x		x
Air Quality			x	x	x	x
Zoning Plan		x	x	x	x	x

Table 2.1: Urban data classes availability on open data portals

Esri Shapefiles. Esri Shapefile [8] is generally a vector data file format for a geographic information system. A single shapefile consists of several files; the minimal set of files necessary for correct data interpretation is DBF, SHP, and PRJ file:

SHP contains the actual geometry,
DBF contains the attributes and indexes,
PRJ describes the coordinate reference system.

In the original specification [8], SHX is also a mandatory file. Still, the file only allows to accelerate the queries and is not required for the data to be displayed correctly. The Shapefile format is well documented, and there is a range of existing importers in various languages.

CityGML and CityJSON. CityGML [9] is an XML-based file format designed to capture the structure and metadata of 3D city models. The aim of the development is to come up with a standardized and sustainable way of storing urban data. There is a JSON-based version of CityGML - CityJSON [10]. Despite minor differences, these two standards are mostly compatible. From the practical point of view, the CityJSON format is much less verbose and thus easily parsable. Both CityGML and CityJSON are open formats; there are converters and importers available .

GeoJSON. Another JSON-based standard is the GeoJSON format. This GeoJSON structure is quite simple; the file can describe a set of primitive geometry objects such as lines, lines, or polygons. Each primitive can have a variable number of properties, which can be completely arbitrary. The GeoJSON file usually describes data such as networks or areas. Compared to CityJSON, GeoJSON is not suitable for representing hierarchies, which usually naturally appear in building geometry and its components. GeoJSON is an open format, it is fairly simple to write an importer for this format, and there are several available, see [11, 12].

DGN and DWG. Both of these formats are proprietary vector formats used by CAD editors, such as AutoCAD. A reader for the DWG format is a part of RealDWG developer suite created by Autodesk [13]. Its license is incompatible with most free software licenses. As a result, no open tools contain an effective importer for the DWG file format. There is no official free DWG SDK; the software for importing and converting DWG is a part of software offered by OpenDesignAlliance (ODA) [14]. Some sort of DWG and DGN loading capabilities are implemented in the GDAL framework [15].

BIM and IFC. Based on [16], Building Information Model (BIM) is the common name for all data relevant to an architectural structure — the plans, costs, changes over time, etc. Industry Foundation Classes (IFC) is a data model designed for architectural, building, and construction data. It is a platform-neutral and open file format. In some states, the IFC has been legally established as the standard exchange format for public project data. Like Shapefiles, IFC is not a single file; there are many available IFC versions, and the development of newer ones is underway. On the other hand, contrary to Shapefiles, each IFC file version should include the same data; the only difference is in the encoding.

2.5 Models Based on Urban Data

Urban data alone is not always a sufficient input for city planning. The main issue is the static nature of the data. According to Ira Winder [17], analytical models presented as interactive simulations are often more suitable for the purposes of planning. The models help answer questions. There is no guarantee the models are correct; complex models can be just as useful as the simple ones.

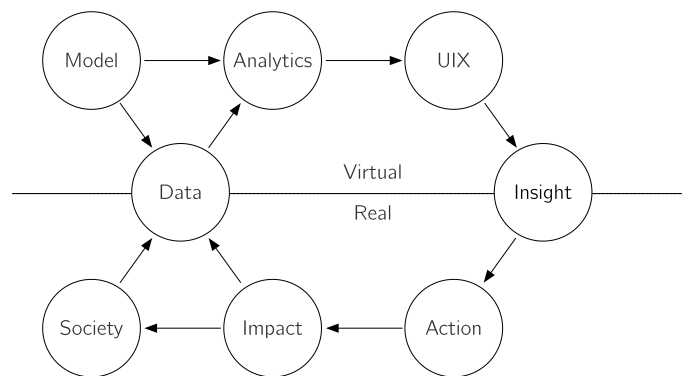


Figure 2.1: Sociotechnical system model [1]

In a talk given by Ira [1], he describes the paradigm of generating insight from data, which is illustrated in figure 2.1. He emphasises the role of the

user interface, which is essential in getting familiar with the actual data and analytics. The generated insight influences further planning and development. The impact of these actions has the potential to generate new data which feeds back into the loop. The step from data to analytics can be rather intensive and involves creating analytical models.

Chapter 3

Existing solutions

The urban data visualization process can be viewed from multiple perspectives. We can consider the actual physical form of the presentation — the medium and the interaction mechanism. On the other hand, it is possible to consider the tools used for the visualization creation, which is more closely related to the standard visualization pipeline. Considering the physical form of the visualization, the examples listed in this chapter can be divided into categories — purely virtual visualizations, and visualizations with physical components.

3.1 Visualization Tools with Physical Components

This section presents a selected list of visualization tools and projects which use interactive physical components in any shape or form. Due to the nature of these tools, the environment where the visualization is presented plays an important role. It has a profound influence on accessibility, interactivity, and the overall user experience. All presented examples share a common scheme — a table with interactive physical components is accompanied by a wide-screen projection presenting the actual visualization.

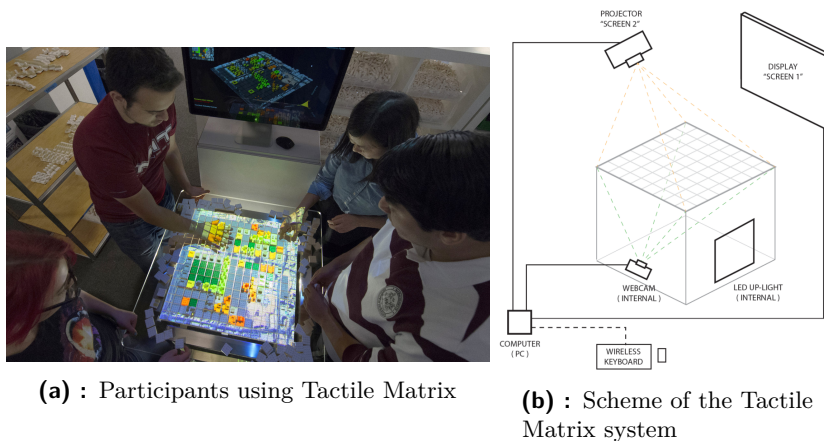


Figure 3.1: Tactile Matrix framework

Tactile Matrix. Ira Winder and Kent Larson have presented a framework called Tactile Matrix [?]. It is a collaborative tool for real-time computation and projection mapping, see figure 3.1. The tool is based on a principle described in section 2.5. The users are presented with a matrix representing city blocks; additional information is conveyed by the visualization projected on top of the matrix. The visualization is driven by an analytical model, which takes the matrix configuration as an input. Lego blocks can be placed into the matrix and the system updates the projection, as the model responds to new inputs. The tool is implemented in Processing, which is a Java-based environment. The SDK for the Tactile Matrix project is available, which also includes the manual to construct the physical components.

■ 3.2 Purely Virtual Visualization Tools

Appendix A

Bibliography

- [1] Ira Winder, MIT Media Lab, Big Data Analytics Tokyo 2017. In: YouTube [online]. 12. 5. 2017 [cit. 10. 9. 2020]. Available from: https://youtu.be/zVmc5_KK5JU.
- [2] Foth, M.; Choi, J. H.-j.; et al. Urban informatics. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011, pp. 1–8.
- [3] Robinson, R.; Rittenbruch, M.; et al. Street computing: Towards an integrated open data application programming interface (API) for cities. *Journal of Urban Technology*, volume 19, no. 2, 2012: pp. 1–23.
- [4] Weiser, M. The Computer for the 21 st Century. *Scientific american*, volume 265, no. 3, 1991: pp. 94–105.
- [5] Zheng, Y.; Liu, Y.; et al. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*, 2011, pp. 89–98.
- [6] Burke, J. A.; Estrin, D.; et al. Participatory sensing. 2006.
- [7] Ross, L. *Invitation to participate in the 5CC group*. [email]. Message to: Lochschmidtová Kateřina Ing. Mgr. lochschmidtova@ipr.praha.eu, Čtyrský Jiří Mgr. Ph.D. ctyroky@ipr.praha.eu. 13. 10. 2020 [cit. 30. 10. 2021]. Personal communication.
- [8] ESRI. ESRI Shapefile technical description. Technical report, 1998.
- [9] Gröger, G.; Kolbe, T. H.; et al. OGC city geography markup language (CityGML) encoding standard. 2012.
- [10] Ledoux, H.; Ogori, K. A.; et al. CityJSON: A compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, volume 4, no. 1, 2019: p. 4.
- [11] CityJSON: Software [online]. CityJSON, 2019. [cit. 30. 8. 2020]. Available from: <https://www.cityjson.org/software/>.

- 14