

PAPER • OPEN ACCESS

Data mining in the context of urban metabolism: A case study of Geneva and Lausanne, Switzerland

To cite this article: N S Wiedmann *et al* 2021 *J. Phys.: Conf. Ser.* **2042** 012020

View the [article online](#) for updates and enhancements.

You may also like

- [City, Space and Publicness: Perceptions and Experiences in The Case of Isparta](#)
Huseyin Gul, Nilufer Negiz and Seda Efe
- [Renewal and Upgrading System For A Sustainable Urban-Rural Housing System Development: Panacea To Accommodation, Employment And Healthcare Issues](#)
L. M Amusan, C.N Akanya, K. A Adeyemo et al.
- [Estimating current and future global urban domestic material consumption](#)
Timothy Malcolm Baynes and Josephine Kaviti Musango



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Data mining in the context of urban metabolism: A case study of Geneva and Lausanne, Switzerland

N S Wiedmann, A Athanassiadis¹ and C R Binder

Laboratory for Human Environment Relations in Urban Systems (HERUS), Institute of Environmental Engineering (IIE), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

1: Author to whom any correspondence should be addressed

Aristide.Athanassiadis@epfl.ch

Abstract. The highest share of the global population lives in cities. The current configuration of the latter requires considerable amounts of resource flows causing the degradation of local and global ecosystems. To face the complexity of these challenges, scientists use the concept of urban metabolism (UM), i.e. measuring urban input and output flows from a systemic perspective. This accounting method results in a large data collection from multiple sources that are often not harmonised. Metabolism of Cities Data Hub is an online platform which facilitates data collection, processing and visualisation in order to extract urban metabolism insights. This work highlights the challenges faced when mining urban metabolism data in the case of Lausanne and Geneva, as well as provides insights on how data could be best used from users and providers. Slight differences between the two case studies, in terms of data accessibility and availability were experienced but the main challenges revolved around data copyright, format and availability. As a conclusion, the used tool can enable harmonisation and standardisation of UM data. As such it could contribute to the use of data mining to streamline the environmental monitoring of cities as well as facilitate the creation of mitigation strategies.

1. Introduction

Cities are the arenas where social, economic, political, cultural activities co-exist and are deeply intertwined. The United Nations estimates that 56% of the world population now lives in urban areas and this share is even higher in Switzerland where it is estimated to approach 74% [1]. The current functioning of urban areas and their associated activities are responsible for the extraction and transformation of resources which in turn result in the degradation of local and global ecosystems but are also directly or indirectly linked to profound societal crises [2]. The complexity of these interlinkages makes it difficult to discern how to deal with, let alone tackle, them. The systemic understanding and analysis of urban areas appears thus as a fundamental piece to explore local and global challenges and propose adequate strategies.

Urban metabolism (UM) is a metaphor that has been used by several academic (sub-)disciplines to study the relationship between urban activities and their relationship to local and global ecosystems from a systemic perspective. One strand of researchers focuses more specifically on the accounting of resource use flows entering, being transformed and or stocked, and released in the form of pollution flows [2]. This environmental accounting procedure uses different accounting methods ranging from



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Material Flow Analysis, Input-Output Analysis or Life Cycle Assessment and can highlight specific flows or economic activities that are responsible for the highest share of environmental pressures.

In most cases, quantification is a result of a large (often structured) data collection exercise, harvesting data through official reports, datasets, companies’ annual reports, grid operators, etc., which entails a wide range of formats and indicators [3]. Yet, most of the data collected are neither harmonised nor easily machine readable to make relevant, detailed, and semi-automated analyses.

To address the above-mentioned challenges, Metabolism of Cities (MoC) [4], an online and open source platform, developed a Data Hub which collects, processes and visualises data in order to best illustrate the insights coming from UM studies. The current paper illustrates the said challenges faced when data mining in the context of UM using Lausanne and Geneva, Switzerland, as case studies and the MoC platform as a tool. As expected, some of the difficulties encountered during this research were mainly linked to data copyright, formats, and availability. The results from this research are two online dashboards which collected all available information to analyse the UM of Lausanne and Geneva with different levels of granularity. The paper will then discuss the relevant insights for data users (researchers and urban policy makers) and data providers (statistical offices, official departments) stemming from the dashboards. Finally, the conclusion section will dive into how and whether the use of the Data Hub could further facilitate data mining for future UM studies.

2. Methods and case studies

This section provides contextual information on the two case studies selected for this research as well as the accounting framework implemented in the Data Hub and used to carry out an UM study.

2.1. Case studies

Independently from the selected method to perform an UM study, namely how resource and pollution flows and material stocks are accounted for, a system boundary, defined both in space and time, needs to be clearly delineated before launching the data collection process. In this study, two spatial scales were chosen for the two case studies: the cities themselves (Lausanne City and Geneva City) and their respective cantons (Vaud and Geneva). The image on the bottom shows contextual information for the year 2018, that provides a whole picture of the selected systems (Figure 1).

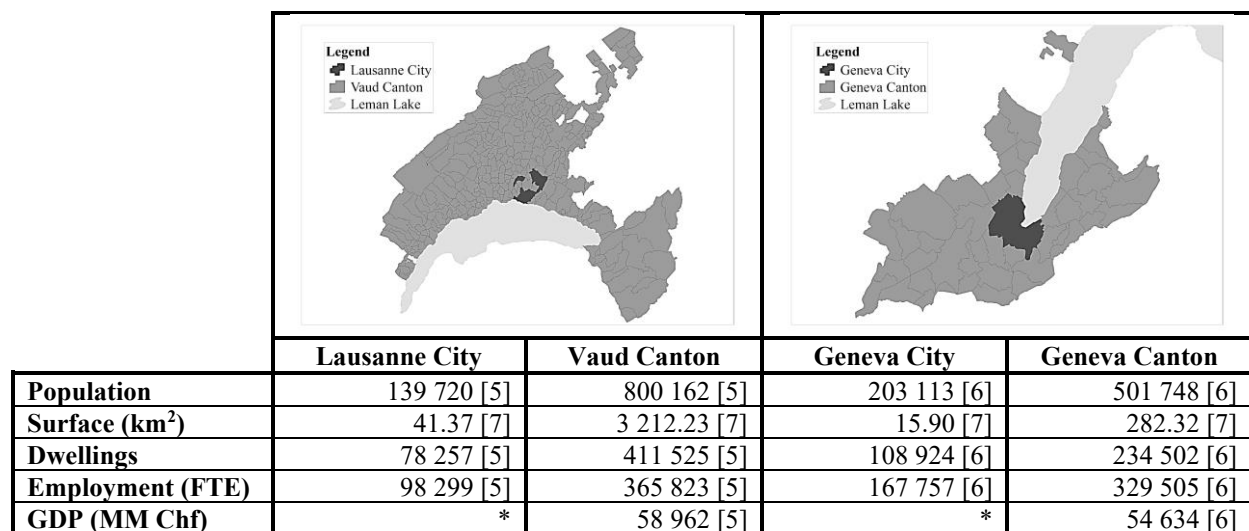


Figure 1: Case studies context in 2018. Sources: [5,6,7].

Boundaries were selected to have a multiscale and comprehensive overview of the territorial systems and their associated activities (from consumption, production, transportation, etc.) considered. Lausanne and Geneva municipalities are at the core of their cantons, hosting a great share of their population and consequently of dwellings, employment, and services. Cantons present the larger administrative boundaries which define major policies and consequent development of the territory.

Regarding time boundaries, as the MoC Data Hub enables to collect datasets covering many years for the same spatial system, a time-series ranging from the oldest to the most recent records was carried out.

2.2. *Metabolism of Cities (MoC) Data Hub*

UM studies measure the resource use and pollution flows as well as material stocks of cities. Several accounting methods (material flow accounting, etc.) are available to collect and analyse data in order to extract different insights. In this paper, we have opted to use the MoC Data Hub to structure our data collection and analysis through the use of layers and sublayers [4].

The Data Hub layers were inspired by Kennedy et al. [8], on top of which were added infrastructural and spatial components. More specifically, the first layer of the Data Hub covers contextual information necessary to define the system boundaries and its core, such as data on population, economic activities, and policies. A second layer regarding biophysical characteristics covers natural resources present within the territory and its climate. The third layer provides an understanding of infrastructures that mobilise and transform flows. The last layer concerns flows and stocks *per se*, i.e. natural extraction, emissions to the environment, imports and exports, as well as the material stocks, i.e. the mass that stays in an urban system for more than a year. To fill in the four layers of the MoC Data Hub, a manual data collection, extraction and formatting was applied. Once raw data was uploaded on the platform (geolocalised information, spreadsheets, reports, articles, etc.), they were then processed using standard templates to become machine readable and interactively visualised in forms of maps or charts.

2.3. *Data Collection*

Data have been harvested to cover information on the city/canton scale following the different layers and sublayers proposed by the Data Hub. To complete these four layers (and their associated sublayers) a wide range of information was uploaded from several sources.

In our case, data was mainly collected from official departments. These included industrial and water services for water and energy consumption and distribution related data [9,10,11]. Municipal and cantonal departments (energy, environment, and others) provided territorial information [12,13]. Data on agricultural extraction were obtained through different individual requests [14,15]. Yet, most of the data came from cantonal statistical offices and the Federal Statistical Office [5,6,7].


3. **Results**

The result from this data mining exercise resulted in the development of two online dashboards, one for each of the territories studied, including all relevant UM data at a city and cantonal scale for Lausanne city/Vaud canton, and Geneva city/Geneva canton. About 200 documents were collected for the two spatial scales of Geneva and 150 for those of Lausanne [4]. Depending on the considered layer and sublayer, a different quantity and quality of information was available. It is also important to mention that while for some layers one dataset is sufficient to cover the topic, in other cases multiple sources of data are necessary to encompass its complexity.

Contextual information was easily accessible from statistical offices and provided a comprehensive overview of the systems considered. Several datasets on population and population structure, on economic activities presenting employment and company characteristics were available. Municipal or cantonal policies are exposed in open access on the city or canton websites.

Biophysical information is mostly available by territorial related departments. In the case of Lausanne and Canton Vaud reports and online maps were mainly available, while in the case of Geneva many open access geolocalised information were available. Vaud canton also disposes of geolocalised information but was not easily accessible. Despite the mix of format and sources, a good understanding of systems' biophysical characteristics became possible.

Figure 2 presents a synthetic overview of available data with a note on the format's quality and on accessibility.



Data Quality	Lausanne City			Vaud Canton			Geneva City			Geneva Canton		
	A	F	OA	A	F	OA	A	F	OA	A	F	OA
1. Context	Boundaries, Economic activities, Population & Policies											
2. Biophysical	Soil, Vegetation, Biodiversity, Water, Climate & Minerals											
3. Infrastructure	Land use, Networks, Services, Manufacturing sites, Transports...											
4. Stocks & Flows	Natural extraction											
	Consumption, budget surveys											
	Stocks: Buildings, Infrastructure											
	Stocks: Vehicles											
	Stocks: Livestock											
	Flows: Energy											
	Flows: Water											
	Flows: Imports & Exports											
	Emissions (air, water, products)											
	Waste flows											

Figure 2: Matrix presenting data collected by availability and quality. A=Availability; F=Format; OA=Open Access.

In the case of the infrastructure layer, the quality and quantity of information harvested varied widely depending on the sublayer considered. Basic information on land use, agricultural exploitation and utilized agricultural area, mining, and lodging, are available in spreadsheets format, mainly at a cantonal scale. Data on water, energy and waste infrastructures are generally reported by services in charge of their distribution and collection in the format of reports. For the infrastructure sublayers, geolocalised information is often available but it is only fully accessible through confidentiality agreements for Vaud canton. On the other hand, a comprehensive overview of other specific types of infrastructure and manufacturing sites was almost impossible to carry out as they are poorly documented by industry associations in form of reports or websites.

Stocks and flows data are principally available from cantonal and federal statistical offices. Data on natural extraction are reported at cantonal scale with different metrics which require further adaptations to reflect quantities. In general, there is a lack of data on industrial and services production and their associated material consumption. Nevertheless, water and energy use is well documented mainly as there are individual meters in buildings and companies. Data on food consumption were available through household budget surveys which are conducted on a larger scale than the city/cantonal one. Stocks of vehicles, buildings, infrastructure, and livestock are generally expressed by their number from statistical offices. Data on flows are generally less accessible. Regarding imports and exports, data are collected considering the exchanges of foreign markets between Switzerland and abroad at a national and cantonal scale but not at a city scale. Imports and exports also lack intra-national trade flow making it hard to construct an overall image. Emissions to air flows were recently developed from official departments, and mainly evaluate greenhouse gas emissions. A first outlook on emission to water was available through water treatment plants reports. Data on emissions from dissipative use of products were not available. Regarding waste production, several datasets were available in open access format, but considering the different indicators used and their respective granularity, it is difficult to obtain a whole picture of this flow.

After data processing, i.e. formatting of data in a machine-readable format to the Data Hub, the platform enables to host and visualise UM data. This consists of interactive maps and datasets that can be consulted online and give first insights to the city/cantonal dynamics, namely main stocks and flows and their evolution throughout the years, as well as their associated infrastructures. The data collection and storage on the Data Hub helps to understand which are the existing data and the most relevant, as well as identifying data gaps. This also enables new users to avoid starting from scratch their data

harvesting as well as develop new avenues of collaboration between researchers. Figure 3 shows UM insights offered by MoC Data Hub.

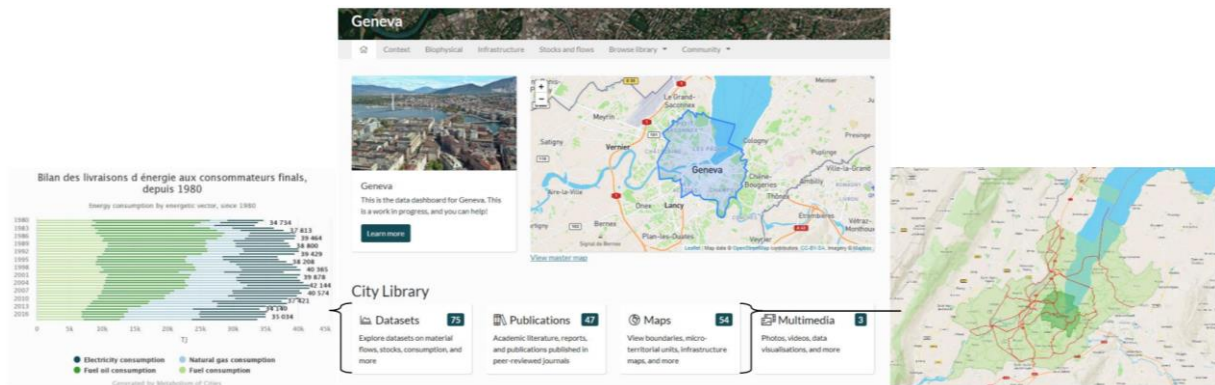


Figure 3: Overview of Geneva’s dashboard and data analysis example. The chart shows energy consumption in Canton Geneva, the map high-voltage lines and fuel pipelines [4].

4. Discussion

During the data collection process and the subsequent elaboration of two online dashboards some challenges emerged. These can generally be divided in three categories: copyright, data format and availability.

Data provided by some official and administrative departments are frequently either copyrighted or the license to use them is not expressed clearly enough. Some specific datasets protected by confidentiality agreements were not uploaded in the open access Data Hub. These above-mentioned challenges not only hinder the accessibility of data but also how easily they can be mined in the future. Whenever this was the case, steps on how to gather the data were documented on the platform. However, some efforts are being made, for example the Swiss Geoportal just expanded the share of its open access data [16]. Nevertheless, the most crucial factor which could contribute in developing data mining techniques in the field of urban metabolism is the harmonisation of data formats. Most data providers use their unique format to publish data (geolocalised informations vs. spreadsheets vs. reports). Spreadsheets is currently the best format to provide data on stocks or flows quantities in order to have a full dataset yearly updated. In the case of biophysical characteristics, infrastructure, and stocks and flows’ layers linked to a specific site, geolocalised information is the best way to provide a dataset because it allows to know exactly which resources are present, and where. Moreover, for each (sub-) layer there are different metrics proposed, which have different degrees of pertinence in UM studies and should sometimes be translated to relevant units to extract insights.

While the MoC Data Hub has provided a set of layers and sublayers to structure UM data collection as well as templates to format data, these have been designed to cater for different formats, nomenclatures, and methods. This flexibility from the Data Hub enabled the crowdsourcing of UM data. This method could also be scaled to countries or larger cities. The number of collected datasets is not proportional to the vastness of a territory. Moreover, in a larger context more pertinent data could be available. Nevertheless, to truly identify patterns and extract new knowledge from the study of the metabolism of several cities, a standardised indicator set, or accounting method might need to be privileged in the future.

5. Conclusion

Cities are the territories where sustainability challenges are created and might get solved. Yet, the sheer magnitude and complexity of these challenges call for systemic methods of analysis. UM is a promising method to account for resource use and pollution emission flows and identify their drivers in order to mitigate the former. Yet, due to the inconsistent format and accessibility of relevant datasets such insights are not yet available. The MoC Data Hub offers a way to centralise collection and analysis of

such data. The current paper underlines a first attempt to carry out UM data mining for Geneva and Lausanne. The major challenges of data mining were underlined and solutions to increase the uptake of data mining in the UM field were outlined. As a conclusion, the MoC Data Hub could provide a solid foundation to streamline environmental monitoring of cities and the identification of mitigation strategies. In the future, considerable efforts need to be made both by data providers and users in order to facilitate this process.

References

- [1] United Nations, Department of Economic and Social Affairs, Population Division 2018 *World urbanization prospects: The 2018 revision* Online Edition [Internet]. [cited 2021 Apr 16]. Available from: <https://population.un.org/wup/Download/>
- [2] Mirelles J, Pagani A, Athanassiadis A and Binder CR 2020 Sustainability issues in urban systems from a metabolic perspective. *Sustainability Assessments of Urban Systems*. Cambridge University Press. p. 29:261.
- [3] Athanassiadis A. 2020 Urban metabolism and open data: Opportunities and challenges for urban resource efficiency. *Open Cities | Open Data*. p. 177:196.
- [4] Metabolism of Cities Data Hub [Internet]. [accessed 2021 Apr 16]. Available from: <https://data.metabolismofcities.org/>
- [5] STATVD. Statistiques cantonales - Etat de Vaud [Internet]. [accessed 2021 Apr 22]. Available from: <https://www.vd.ch/themes/etat-droit-finances/statistique/statistiques-par-domaine/>
- [6] OCSTAT. Statistiques cantonales - République et canton de Genève [Internet]. [accessed 2021 Apr 22]. Available from: <https://www.ge.ch/statistique/domaines/welcome.asp>
- [7] BFS. Federal Statistical Office [Internet]. [accessed 2021 Apr 22]. Available from: <https://www.bfs.admin.ch/bfs/en/home.html>
- [8] Kennedy C, Stewart ID, Ibrahim N, Facchini A and Mele R 2014 Developing a multi-layered indicator set for urban metabolism studies in megacities. *Ecol Indic.* **47** p. 7:15.
- [9] Ville de Lausanne. Service de l'eau [Internet]. Site officiel de la Ville de Lausanne. [accessed 2021 Apr 22]. Available from: <https://www.lausanne.ch/officiel/administration/securite-et-economie/eau.html>
- [10] Ville de Lausanne. Services industriels Lausannois (SiL) [Internet]. Site officiel de la Ville de Lausanne. [accessed 2021 Apr 22]. Available from: <https://www.lausanne.ch/vie-pratique/energies-et-eau/services-industriels.html>
- [11] Services Industriels Genevois - SIG [Internet]. [accessed 2021 Apr 22]. Available from: <https://ww2.sig-ge.ch/>
- [12] Etat de Vaud. ASIT VD [Internet]. [cited 2021 Apr 22]. Available from: <https://www.asitvd.ch/>
- [13] SITG. Le territoire genevois à la carte [Internet]. [accessed 2021 Apr 22]. Available from: <https://ge.ch/sitg/>
- [14] SZG CCM CSO. Centrale Suisse de la culture maraîchère et des cultures spéciales [Internet]. [accessed 2021 Apr 23]. Available from: <https://www.szg.ch/it/>
- [15] Union suisse des paysans. Agristat : statistique de l'agriculture suisse [Internet]. [accessed 2021 Apr 23]. Available from: <https://www.sbv-usp.ch/fr/service/agristat-statistique-de-lagriculture-suisse/>
- [16] Swiss Geoportal [Internet]. geo.admin.ch. [cited 2021 Jun 2]. Available from: <https://map.geo.admin.ch>