



## **Towards spatial integration of qualitative data for urban transformation – challenges with automated geovisualization of perception of urban places**

Downloaded from: <https://research.chalmers.se>, 2024-06-20 09:12 UTC

Citation for the original published paper (version of record):

Thuvander, L., Latino, F., Zangelidis, C. et al (2020). Towards spatial integration of qualitative data for urban transformation – challenges with automated geovisualization of perception of urban places. IOP Conference Series: Earth and Environmental Science, 588(5). <http://dx.doi.org/10.1088/1755-1315/588/5/052041>

N.B. When citing this work, cite the original published paper.

PAPER • OPEN ACCESS

## Towards spatial integration of qualitative data for urban transformation – challenges with automated geovisualization of perception of urban places

To cite this article: L Thuvander *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **588** 052041

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

# 239th ECS Meeting

with the 18th International Meeting on Chemical Sensors (IMCS)

**ABSTRACT DEADLINE: DECEMBER 4, 2020**



May 30-June 3, 2021

**SUBMIT NOW →**

# Towards spatial integration of qualitative data for urban transformation – challenges with automated geovisualization of perception of urban places

L Thuvander<sup>1</sup>, F Latino<sup>1</sup>, C Zangelidis<sup>2</sup>, M Adelfio<sup>1</sup>, V Naserentin<sup>2,3</sup>, A Logg<sup>2</sup>

<sup>1</sup>Department of Architecture and Civil Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sven Hultins gata 6, Sweden

<sup>2</sup>Department of Mathematical Sciences, Chalmers University of Technology, 412 96 Gothenburg, Chalmers tvärgata 3, Sweden

<sup>3</sup>Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Corresponding author: liane.thuvander@chalmers.se

**Abstract.** Urban planning needs to face and integrate ecological, social, and economic aspects of city living. So far, attempts to integrate different urban simulation models into one technical platform have focused on quantitative data. The aim of this paper is to present the preparation for an automated method to spatially integrate and visualize interview-based qualitative data on the perception of urban places into a virtual platform. The Gothenburg suburb of Hammarkullen is used as a case study. Two CAQDAS software, NVivo and Atlas.ti, were tested. In both software, locations and urban qualities were coded and clustered. Visualization strategies such as information tree structures, geo-coded quotations, spatial word clouds, linked to 2D maps and 3D environments were developed. Results identify the challenges to overcome and show limitations of the software in terms of creating graphs and relationships as well as automated geocoding of data to maps. The project offers a step towards the integration of qualitative (social) data into digital environments that can be scaled up. By that, it contributes with a new dimension of analysis of urban environments which is necessary for sustainable transformation of cities.

## 1. Introduction

Urban planning needs to deal with multiple dimensions of urban living and contributes to shaping urban places. The latter can be examined objectively via quantitative data sources or subjectively according to individual perceptions [1]. The perception [2] of urban spaces is embroiled in two dichotomies. First, the dichotomy between its physical and virtual dimension, synthesized in Habermas' idea of public sphere, a "virtual or imaginary community which does not necessarily exist in any identifiable space" [3:217]. Within such debate, local context plays a key role in shaping urban spaces as "the built environment is rich with meaning", forged by local communities [4:46]. Second, urban places can be analyzed through a mix of quantitative and qualitative methods using "intersubjectivity" [5:71] as a research approach that collects multi-stakeholder research inputs embracing a diversity of perceptions. The focus on the integration of qualitative and inter-subjective data supports the UN Sustainable Development Goals (SDG) 11 "Sustainable cities and communities" from a community point of view.



There are growing efforts to integrate qualitative data into GIS and digital platforms [6], to some extent using automated procedures (e.g. [7]). Few scholars are using automated procedures and even less are managing to obtain a semantic and ontological integration that goes beyond a mere technical interlinking of different software - e.g. GIS and Computer-Assisted Qualitative Data Analysis Software (CAQDAS). The obtained results are often limited to quantitative visualizations of qualitative content. The challenge is to achieve an effective, meaningful, and automated integration of qualitative data [8].

The VirtualCity@Chalmers project (<https://virtualcity.chalmers.se/>) carried out at Chalmers University of Technology, Sweden, integrates different urban simulation models into one technical platform. So far only quantitative simulation models are integrated, showing the need for incorporating the missing qualitative data to support decision makers.

The aim of this study is to test how current CAQDAS can enhance the spatial integration of social data in digitalization processes. Such a techno-innovative focus supports the UN SDG 9 “Industry innovation and infrastructure”. A long-term goal is to develop an automated method to spatially integrate and visualize qualitative data on the perception of urban places into a virtual platform. The following research questions were explored:

- How to keep, present, and visualize feelings and other intangible social data in a spatial, digitalization process?
- Do existing software for qualitative data analysis, such as NVivo or ATLAS.ti, enhance spatial integration of social data in digitalization processes?
- What are the potentials for automation of the process and integration of qualitative and quantitative data?

The study was carried out as pilot study limited in time and focused on a test case area. The virtual 3D platform developed in the VirtualCity@Chalmers project was used as framework. The platform is still in closed beta phase and now part of the Digital Twin Cities Centre, a VINNOVA competence center hosted by Chalmers (2020-2025).

## **2. The test case area Hammarkullen**

For the study, the district Hammarkullen was used, a suburb located about 10 km northeast of the city of Gothenburg, Sweden, constructed during the 1960s and 70s, as part of the so-called Million Home Programme era, to respond to urgent housing needs and housing degradation in the city center. The area is predominantly free of cars and consists of mainly large-scale housing, but also lower buildings, detached houses and villas. Today about 8,200 people are living in Hammarkullen, about 57% were born abroad [9] and 84 nationalities are represented, speaking over one hundred languages. Chalmers has carried out research projects and architectural courses in the area for more than 10 years, often working together with local stakeholders. Thus, the area was highly relevant for our pilot study.

## **3. Method**

The project was carried out as a pilot study to explore future potentials and to unveil challenges for the development of methods for spatial integration of qualitative data for automated as well as geovisualization of perception of urban places. The pilot study was limited in time (4 months). The emphasis was therefore on investigations of available tools and software and their capacities rather than on collecting a large sample of empirical data.

The project started with a desktop literature review focused on social data representation in digital environments and the perception of living, working and meeting places. Then qualitative data were collected through 12 short semi-structured interviews with residents and working people of Hammarkullen. Thereafter, the usability of the two software CAQDAS NVivo and Atlas.ti was evaluated, other data sources related, and the Virtual Cities platform prepared for integration.

For the interviews, people were randomly asked on the street, playground and in shops if they were willing to participate. Criteria were to find people with different profiles of ages and gender at the different places. In total, 10 females and 5 males with ages ranging from approximately 20 to 60 agreed; two interviewees were below 18 but were accompanied by an adult who also agreed to be interviewed.

The interviews were carried out in English, on a weekday afternoon in spring, during a 5 hour period, and were 3 to 12 minutes long. The voice-recorded interviews were transcribed (manually and automated in a test). For the automated transcriptions, three software were tested: Google cloud speech-to-text, Trint, and Happyscribe, but the results were not satisfactory as the transcriptions were not accurate enough. This might be due to the quality of the audio file and/or the pronunciation of the interviewees whose first language was not English, or it was a software limitation. The transcriptions were thereafter integrated, coded and clustered in NVivo and Atlas.ti.

The questions we asked concerned living, working and meeting in Hammarkullen. All interviewees were asked if they live in Hammarkullen, how they perceive Hammarkullen as a place to live/work/meet people, and if they feel at home in Hammarkullen. Other questions were adapted depending on the location of the interview. Regarding places to live, we asked the interviewees who lived in Hammarkullen where they live and how they feel about their home and the surroundings. Interviewees who did not live in Hammarkullen were asked if they have any opinion of specific places in the area and if yes to point them out on the map. Regarding working places in Hammarkullen, we asked for the reasons why they like or don't like working there (shop/area). Regarding meeting places in Hammarkullen, we asked for examples of places they meet and reasons why they meet there. In accordance with the questions, during the interviews, places were marked on a paper map. For the interview location, interviewees home or work places, and the mentioned places, the coordinates were noted. The coordinates of the locations of interest were specified at a later stage by the researchers and were informed by the markings on the paper maps during the interviews. While walking around in the area and conducting the interviews, the interviewers also made observations how and where people moved and what places they used.

The coding and categorization of the transcripts focused on the locations the interviewees mentioned, the qualities they described, and the perception and interpretation of them (positive/neutral/negative). For the coding of the location as a place to work, live or meet, an open coding was applied. For the following categorization of related qualities and specification of the research questions used for the QDA, i.e., the perception of places for living, working, and meeting in Hammarkullen, a selective coding was adopted.

For the places identified by the interviewees, we also examined if places had architectural qualities. Qualitative data from interviews were combined with other qualitative data sources, e.g., site observations and architectural sketches, and quantitative, statistics and architectural physical attributes.

Attempts were made to automate parts of the processes such as the development of a machine learning algorithm able to predict and classify documents into corresponding nodes. The model works generally well with a large dataset since larger training dataset produce more accurate results. However, due to the small size of our dataset, we were not able to get accurate, automated classifications. Figure 1 illustrates the workflow of our study.

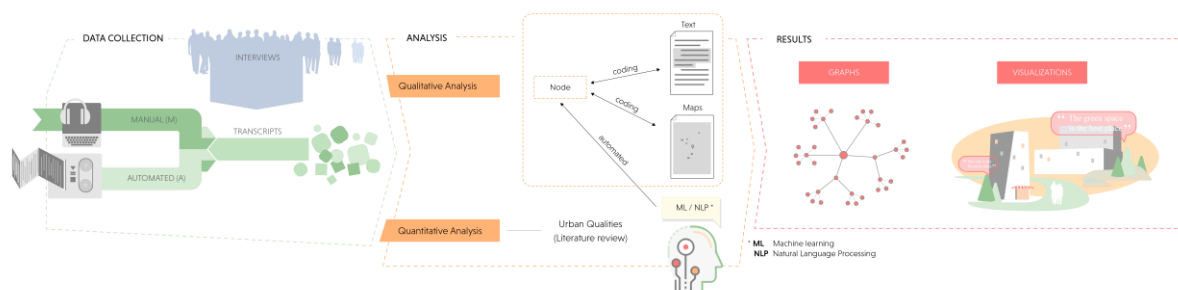


Figure 1. Workflow of the study.

Examples of investigated visualization strategies were both non-spatial graphs such as information tree structures or word clouds and spatial visualization of, for example, (geo-) coded places, quotations and interpreted perception categories, linked to 2D maps and 3D environments.

First efforts were made to integrate results into the VirtualCity@Chalmers platform and to visualize complex social data without losing too much descriptive information. The needed 3D assets for the area were reconstructed from the Swedish cadaster (Lantmäteriet) datasets using a method that procedurally creates watertight volume meshes appropriate for both simulations and visualization tasks (Figure 2).

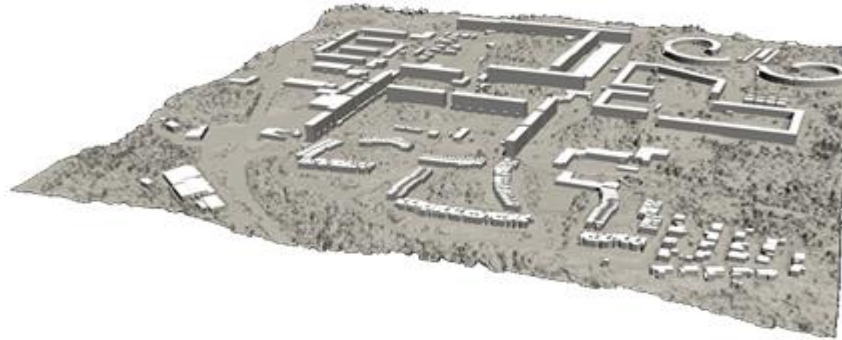


Figure 2. Procedurally generated 3D mesh for the test area Hammarkullen in Gothenburg.

#### 4. Results

Results concern the use and evaluation of the tested CAQDAS and the integration of quantitative measures and qualitative descriptors.

##### 4.1 *The tested software - NVivo and Atlasti*

Both NVivo and Atlas.ti are analysis software for coding or tagging qualitative data such as text based or audiovisual sources. A code is a short word or phrase that symbolically assigns a summative or evocative attribute for a portion of language-based or audiovisual data. The purpose of the software is to divide the data into manageable segments and to have quick access to them for different kinds of analysis. In our case, our data were the imported from transcribed documents of the interviews. We followed the same methodology with importing our data and manually coding them. When going through the data and highlighting interesting and relevant parts of text, codes were assigned. This process had to be done entirely manually as the software is not reading and coding the data automatically. The result was a coding scheme that we used to cluster and visualize our data by generating word clouds or tree structures.

There are many similarities between NVivo and Atlas.ti. but regarding visualization of clusters and relationships, Atlas.ti provides more options. Still, none of the software were sufficiently sophisticated for our purpose. In both cases, a lot of manual work was required, and it was not possible to achieve an automated method for either coding or spatial linking to maps. NVivo didn't work at all for representing different relationships, perception, clustering of qualitative data and keeping quotes. Atlas.ti provided some of the options after manual coding. Figure 3 presents examples of cluster diagram from NVivo and Figure 4 provides examples of a cluster diagram for the software Atlas.ti with quotes of urban qualities of places and the interpreted perception of the places. Figure 5 illustrates a cluster diagram with quotes and links to maps.

##### 4.2 *People-based and place-based urban qualities analysed through mixed quali-quantitative data*

People-based and place-based urban qualities can be described through quantitative indicators and qualitative attributes or descriptors. Considering the value of this paper as a pilot-tested framework for integration of qualitative and quantitative data, it is not meant to provide an exhaustive list of quantitative measures and qualitative descriptors, but rather to identify some examples of such indicators and descriptions whose integration in the framework can be tested. Sources of information are diverse.

Identified social indicators are coming from statistics such as the entropy index as a measure of nationality group diversity [10]. Literature mixed with site observations provided, among others, descriptors of: a) architectural qualities such as hand railings, openings in the facades [11]; b) people-focused qualities of place such as co-presence of people as an expression of density and attractiveness, good visibility for an improved sense of security [12], presence of curbs, edges and pillars giving a sense of privacy [13]; urban comfort related qualities such as the presence of trees [14] or comfortable feeling of air temperature [15]. Some of the qualitative descriptors are also possible to be quantified into measures e.g. number of trees or air temperature.

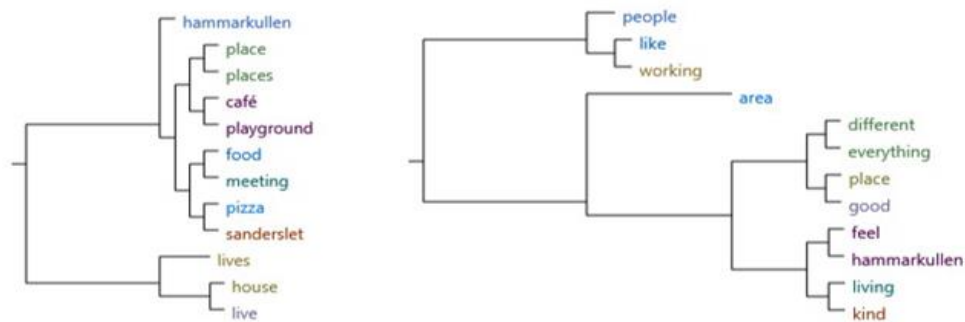


Figure 3. Location node (left) and quality node (right). The location node represents the 12 most frequent words, minimum length in letters: 4 Exact matches (e.g. “talk” in cluster analysis). The quality node represents the 12 most frequent words, minimum length in letters 4 Stemmed words (e.g. “talking” in Cluster analysis)

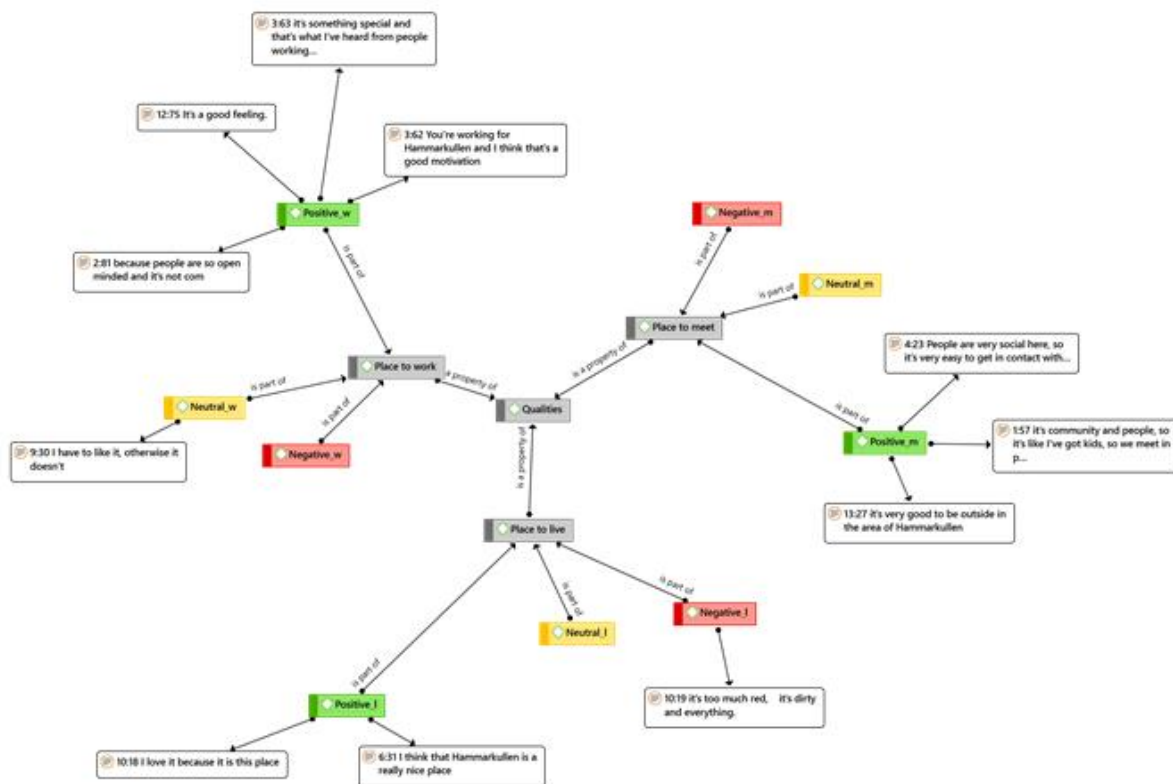


Figure 4. Quality node created in Atlas.ti in 3 steps with quotes from the interviews. 1) creating Nodes, 2) defining relations 3) manual highlights in the text. Green = positive, yellow = neutral, and red = negative perception of place.

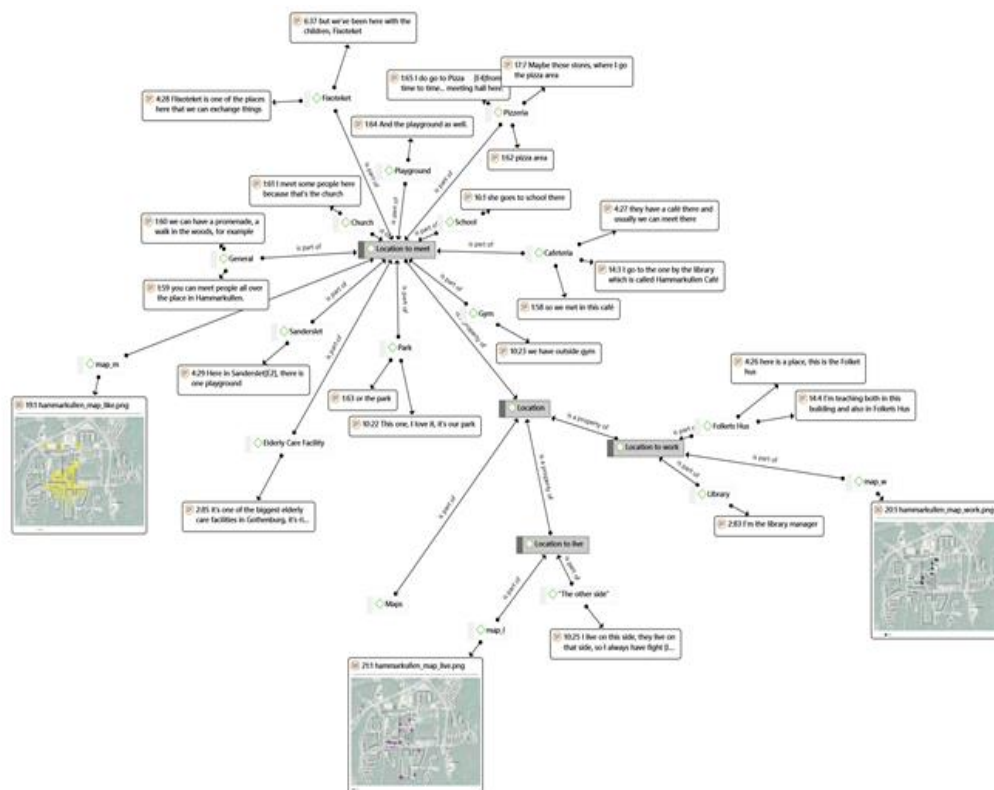


Figure 5. Locations with quotes and links to a map using Atlas.ti.

## 5. Discussion

### 5.1. Challenges with interviews

While the interviewees were able to understand and answer the questions we asked, they were not native English speakers. This might have affected the vocabulary they used to answer the questions and the amount of the descriptive terms. Also, the interviewees used quite often Swedish words to convey what they wanted to say and mentioned specific places in Hammarkullen. Familiarity with the place and the neighbourhood and with Swedish language was necessary to transcribe the interviews correctly. However, the manual transcriptions were carried out by a third party who was not familiar with the places mentioned, and who did not speak Swedish. For this reason, the transcripts presented some inaccuracies that were addressed by the researchers who were familiar with both the Swedish language and know the area and therefore were able to recognize the mentioned places.

Body language and facial expressions were helpful in understanding the interviewees, specifically for coding the answers of the perceived qualities as positive, negative or neutral. As a result, interviews were clearer for the interviewers than for team members who only listened to the audio recordings. Along with the tone of voice, body language and facial expressions provided a layer of information that help when coding the interviews. Thus, when comparing live interviews to audio recordings and transcriptions, in the latter one(s) information is missing. One solution could be to code the audio files and only to transcribe the coded parts later for the visual analysis. Technically, even voice quotations could be kept and integrated in the analysis and geo-coded visualization. However, then special attention needs to be paid to data protection issues.

Another issue is the reliability of the interviews and trust. The interviews were short; people were approached either outdoors and indoors, in public spaces and business premises such as a grocery shops; and the interviewed people were guaranteed anonymity. Although most of the people approached agreed to be interviewed and seemed honest in their answers, establishing trust was one of the challenges encountered by the interviewers, a typical challenge for this kind of interviews. This was especially true



for the people working in Hammarkullen and interviewed at their workplace. As this study was a pilot study with focus on development of the methodology, these challenges could be disregarded. But in a larger study with focus on exploring qualities using this methodology, the considerations need to be addressed.

Automated transcription still encounters limitations and still require manual work. For this project human transcription was preferred. The dataset was limited but a full-scale study should be broader and include more indicators.

### *5.2. Challenges related to the usability of the software and automatization*

The coding was based on wording but remains subjective and informed by the interviewers' perception of the interviewees. Our dataset consisted of unstructured data which is difficult to interpret by a machine learning algorithm and natural language processing. Some challenges are breaking the sentences, deriving context from a discussion, extracting semantic meanings, and accurate polarity detection (negative, neutral, positive). The coding of the interviews in relation to the perception of the places is subjective and connected to interviewers' perception of the interviewees, i.e., if a place is perceived as neutral or positive is difficult to interpret, especially if you do not have the whole sentence, which is also a general challenge in AI language processing. Regarding the usability of NVivo and Atlas.ti to cluster the sentences related to perceptions, both CAQDASs had limitations. Another crucial point is that still a lot of manual work is necessary for the (geo-)coding.

From a graphical point of view, diagrams are easier to build and customize in Atlas.ti than in NVivo. Atlas.ti is more flexible as it allows manual positioning of elements, which makes it easier to establish hierarchy on a visual level. NVivo doesn't allow this flexibility.

In order to support an automated spatial linkage and integration of mentioned places in interviews, future development should integrate digital mapping from the beginning already during the interviews, for example linking CAQDAS with digital mapping tools like KoboToolbox or Maptionnaire. This will require a new way of carrying out interviews.

Due to the obstacles with the software and the resulting lack of time, the 3D model could not be proven in the current study, but we plan to revisit the use of the 3D model in a future study in combination with a larger dataset.

## **6. Conclusion**

A pilot study exploring potentials for the development of a new method for integration of qualitative (social) data into digital environments has been presented evaluating the two Computer-Assisted Qualitative Data Analysis Software NVivo and Atlas.ti. The study has shown that it is possible to create nodes and clusters for perceptions but at the same time the tested software had unexpected, limited potentials regarding automated clustering, visualization and spatial integration. A spatial linking is possible, but an automated spatial integration is still lacking.

Next steps will be to develop automated processes stepwise (interview situation, transcription, coding, spatial integration, machine learning) and to realize a proof of concept for an automated integration of qualitative (social) data into virtual (2D/3D) cities environments. Finally, the project has contributed with a so far under-researched challenge-driven perspective on the integration of qualitative data in an urban context. When fully developed, the methodology can be scaled up and applied in other projects and, by that, contribute with a new dimension of description and analysis of urban environments is necessary for more sustainable transformation of cities and closely linked to the realization of SDG 9 "Industry innovation and infrastructure" but also to the SDG 11 "Sustainable cities and communities".

## **Acknowledgements**

This work was funded by the Area of Advance Information and Communication Technologies as a seed money project and is connected to the project VirtualCity@Chalmers and the Digital Twin Cities Centre at Chalmers University of Technology.

## References

- [1] Haslauer E, Delmelle EC, Keul A, et al 2015 Comparing Subjective and Objective Quality of Life Criteria: A Case Study of Green Space and Public Transport in Vienna, Austria. *Social Indicators Research*. 124 (3) 911-927. <https://doi.org/10.1007/s11205-014-0810-8>
- [2] Lefebvre H 1991 [1974] *The Production of Space*. Oxford: Blackwell.
- [3] Sealey F 2011 *Global Public Square*, Chipmunkpublishing ltd.
- [4] Matthews P 2004 Time, Belonging and Development: A Challenge for Participation and Research. In Nick Gallent N, Ciaffi D (eds.) *Community Action and Planning: Contexts, Drivers and Outcomes*. Policy Press, Bristol.
- [5] Morgan DL 2007 Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods, *Journal of Mixed Methods Research*, 1(1), 48-76.
- [6] Fielding N and Cisneros-Puebla C 2009 CAQDAS-GIS convergence: Toward a new integrated mixed method research practice. *Journal of Mixed Methods Research*, 3(4), 349–370.
- [7] Kyttä M, Broberg A, Haybatollahi M and Schmidt-Thomé K 2016 Urban happiness: context-sensitive study of the social sustainability of urban settings. *Environment and Planning B: Planning and Design*, 43(1), 34-57.
- [8] Adelfio M, Kain J-K, Stenberg J and Thuvander L 2019 GISualization: visualized integration of multiple types of data for knowledge co-production, *Geografisk Tidsskrift-Danish Journal of Geography*, 119:2, 163-184, DOI: 10.1080/00167223.2019.1605301
- [9] City of Gothenburg 2016 Göteborgsbladet – områdesfakta (Area Data). <http://statistik.goteborg.se/Global/Faktablad/G%C3%B6teborgsblad/GbgBlad2016/G%C3%B6teborgsblad%202016-%20Folkh%C3%A4lsoomr%C3%A5den.pdf>
- [10] Börjeson L 2018) *Diversity and Segregation in Sweden*, Swedish Union of Tenants
- [11] Netto V., Vargas JC and de Saboya RT 2019 The social effects of architecture: Built form and social sustainability. In *Urban Social Sustainability* pp. 125-148 Routledge.
- [12] Whyte WH 1980 *The Social Life of Small Urban Spaces*. Washington, D.C.: Conservation Foundation
- [13] Gehl J 1987 *Life between Buildings: Using Public Space*. Washiton D. C.: Island press.
- [14] Gerstenberg T and Hofmann M 2016 Perception and preference of trees: A psychological contribution to tree species selection in urban areas. *Urban forestry and urban greening*. Elsevier GmbH.
- [15] Jeong M, Park S, Song G 2015 Comparison of human thermal responses between the urban forest area and the central building district in Seoul, Korea. *Urban Forestry and Urban Greening* Elsevier GmbH.