Multi-criteria aggregation for sensitive parcel-based census data

S. Indermühle1, P. Laube1, M. Geilhausen1, T. Zwicker2 3

1 Zurich University of Applied Sciences ZHAW
Grietental, 8820 Wädenswil, Switzerland
(indu;patrick.laube;martin.geilhausen)@zhaw.ch

2 tsquare gmbh
Arbergstr. 1, 8405 Winterthur, Switzerland
tzwicker@tsquare.ch

3 City council Männedorf
Bahnhofstrasse 10, 8708 Männedorf, Switzerland

Problem Statement
Modern urban planning considers various socio-economic issues and hence requires fine-grained, potentially sensitive spatial data, raising the disclosure dilemma: Whilst it is imperative to aggregate fine-grained sensitive records large enough to avoid disclosure, aggregation blurs the information, potentially hindering insights about spatially explicit relationships. MASC is an aggregation algorithm for sensitive data, balancing five aggregation criteria.

Validation and Conclusions
The MASC aggregation procedure was evaluated through qualitative plausibility tests with local decision makers. In planning workshops, the stakeholders explicitly noticed that MASC can uncover the fine-grained socio-economic structures whilst safeguarding confidentiality regulations. Manual adjustments were only suggested for a very small number of parcels (around 1% of all parcels).

Given its weighted multi-criteria character, MASC empowers the stakeholders to experiment with different planning scenarios adhering to different priorities (different weights).

References

Acknowledgements
This research was supported by Swiss Commission for Technology and Innovation CTI, project 12910.1 PFES-ES, “modularCity – Software zur Unterstützung nachhaltiger Stadtplanung” and the municipality of Langenthal, for his support.

Case Study
The algorithm is validated using two case studies, Langenthal and Männedorf. The most effective parameter for crystallizing the important socio-economic structure has shown to be resident density (Fig. 2b). Even without any additional information e.g. building type, a high weight of resident density results in zones of the same residential type. By contrast, when accentuating the building regulation parameter (Fig. 2c), the zones are aggregated within respective building zones.

Figure 1
MASC aggregation: (a) all reference areas that are not anonymous (in dark grey) are identified; (b) a first reference area is selected; (c) the difference value with each neighbor is calculated; (d) the reference area is merged with the best match. The resulting reference area is not yet anonymous; (e) the next reference area is selected; (f) the difference value of each neighbor is calculated; (g) after merging with the best match the reference area is anonymous; (h) the 10 anonymous regions resulting from the algorithm.

Figure 2
Different aggregation scenarios, shown for the case study of Langenthal. n was set to 30 residents (a) initial parcels that are used as the input; (b) result based on resident density only; (c) result based on floor area ratio only; (d) results based on geometric distances only; (e) result based on the following exemplary weights: density 1, floor area ratio 0.5, center distances 2.3, polygon distances 2.3, anonymity penalty 2; (f) result based on the following exemplary weights: density 1, floor area ratio 1, center distances 2, polygon distances 0, anonymity penalty 1.