INTEGRATING GEOSPATIAL DATASETS FOR URBAN STRUCTURE ASSESSMENT IN HUMANITARIAN ACTION

Barbara Riedler[1], Stefan Lang[1]

[1]Christian Doppler Laboratory GEOHUM, Department of Geoinformatics, University of Salzburg, Austria
(barbara.riedler, stefan.lang)@plus.ac.at

Commission IV, WG IV/10

KEY WORDS: data integration, data quality dimensions, SDG 11, Earth observation, building footprints, urban planning, informal settlements, humanitarian applications

ABSTRACT:

More than half of the world-population lives in urban areas, with more than 1 billion people lacking basic services and infrastructure. Spatially targeted, data-driven policies are crucial for sustainable urban planning to improve these situations and increase the resilience. Earth observation (EO) can support the process of achieving the SDGs, in particular SDG 11. Aiming at such high-level targets requires a multi-source data environment, defining and extracting suitable EO-based indicators and linking them with socio-economic or environmental data. When embedded in the context of humanitarian response, where physical access to regions is often limited while at the same time, insights on several scales of intervention are key to rapid decisions, the integration of (potentially) heterogeneous datasets requires adequate data assimilation strategies and a good understanding of data quality. This paper investigates the usability of datasets regarding technical and organisational aspects from an application-driven point of view. We suggest a protocol considering various quality dimensions to evaluate via scoring the fitness of multi-source geospatial datasets to integration. The aim is to provide a general orientation towards data assimilability in the context of deriving higher-level indicators, while specific constraints and the need to relativize may occur for concrete use case.

1. INTRODUCTION

1.1 EO and sustainable city development

Currently more than 55% of the world-population (around 4 billion people) live in urban areas; projections show that urbanization will increase in the coming years (World Bank Group, 2021). This requires sustainable planning of city development to increase the resilience of citizens, especially as around a quarter of the world-wide urban population live in slums, informal settlements or inadequate housing, lacking basic services and infrastructure (United Nations, 2021).

Advantages of EO data like high temporal availability, area-wide coverage including remote and inaccessible areas, objectivity, accuracy and reliability and increasing democratization of data, are recognized as key benefits for monitoring and achieving SDGs in general and especially relevant for *SDG 11 Sustainable City Development* (O'Connor et al., 2020; Paganini et al., 2018). This is of particular importance in areas difficult to access and for large, fast-growing cities with urban sprawl.

In the urban context EO data are widely used for sustainable urban development (Prakash et al., 2020) and to map slums, informal settlements or deprived areas (Kuffer et al., 2016; Kuffer et al., 2020). The latter approach is mainly based on building morphology (such as area, shape, height, orientation) and physical characteristics of the near surrounding such as building patterns (Jochem and Tatem, 2021; Kuffer et al., 2016; Taubenböck et al., 2018). These techniques requires reliable building footprints, and thus the need for a comparable evaluation of underlying data quality and suitability is evident.

1.2 Multi-source data integration

Experience with the extraction of single proxies that can additionally contribute to the definition of settlement structure like height information with a higher resolution (Krauß et al., 2019), urban greenness (Kothencz et al., 2018) or distance-based spatial analyses (Hofmann et al., 2015) was gained in various studies. Moving a step ahead towards data integration in remote sensing applications, e.g. urban green classes can be weighted by survey-based preferences of citizens to resemble green valuation (Lang, 2018). In reality, the situation is much more complex. Multiple combined social and environmental factors may affect liveability, wellbeing and deprivation, such as the risk of natural disasters, exposure to diseases, environmental pollution or barriers to services (Abascal et al., 2022).

Operationalising such multi-dimensional concepts and deriving suitable spatial indicators for an integrated assessment requires a multi-source data environment, where EO data are linked with other relevant socio-economic and environmental data. This necessitates techniques to integrate and assimilate data varying in scale, type of measurement, spatio-temporal resolution and extent and combining them meaningfully to set results in a broader social context. Integration of different data through a reliable data assimilation strategy can address complex problems through the identification and description of settlement structure or population distribution models in urban areas to allow a better estimation of where, which and how many people live. This can be a valuable basis for urban infrastructure development and/or logistical planning in humanitarian action, which is of special importance in the current pandemic situation, where the design of vaccination campaigns is more essential than ever.

This paper suggests the usability and assimilability of relevant datasets in the urban context from an application driven point of view. We suggest a systematic approach to evaluate their potential for the analysis or study of complex urban features in a purpose-driven manner (i.e., regarding scale of observation / intervention, as well as application-specific). The following aspects we take into consideration when assessing potential datasets and scoring them accordingly: temporal and spatial resolution, precision, coverage, consistency, completeness, reliability, accessibility and trust of source. Datasets are evaluated based on a fixed scheme and can be ranked according to their scoring and suitability under these aspects. The aim is to provide a general orientation towards usability in the context of higher-level indicators for humanitarian action in an urban environment, e.g. for describing urban structure, while specific constraints and the need to relativize may occur in dependence of the concrete use case.

2. DATA SELECTION

2.1 Data sources for humanitarian action

Here we broadly differentiate between three types of freely available data for the practical use in the urban and humanitarian context:

(i) EO satellite data (primary, pre-processed), e.g. Copernicus Sentinel missions, Open Topography, as well as systematic in-situ measurements

(ii) volunteered geographical information (VGI) or professionally collected data following a certain community-agreed standard, e.g. OSM, Missing Maps, HDX Humanitarian data and

(iii) modelled, derived or interpreted data, e.g. UNOSAT flood maps, Copernicus information services.

A fourth group may arise from semi-voluntary data collection such as call records data (CRD) or georeferenced tweets, which are not primarily collected for mapping purpose. Due to the it´s specific nature and requirements for analyses, this group of data is not explicitly considered in the following (as yet).

In practice, very often multiple data sources need to considered and integrated to support humanitarian action on the ground. Fig.1 illustrates various input data sources - primary data, which undergo interpretation and analysis and additional non-EO data sources towards highly integrated information products relevant in the process of geospatial data integration. The issue of data quality applies to all levels of processing, in the following we focus on the issue of data integration towards an integrated assessment, e.g. by constructing composite indicators. Conducting such advanced spatial analyses like the construction of meaningful spatially aggregated indicators, may be guided by the following steps:

1) Assessment of relevance: which data, indicators, information is needed to answer a specific problem or can support a decision making process
2) Harvesting of existing data: search of collections, data catalogues, interfaces to data portals
3) Information production: derivation and analysis of missing, but needed information (includes classification, spatial analyses, interpolation)
4) Quality control: assessment of usability for all relevant data sources
5) Data harmonization: procedures to make relevant heterogeneous data sources suitable to integrate and assimilate (includes thematic and spatial aspects)
6) Data aggregation: building of in integrated indices if needed (includes statistical analyses, weighting, composite indicators)
7) Spatial integration: integrating relevant information on the scale of need

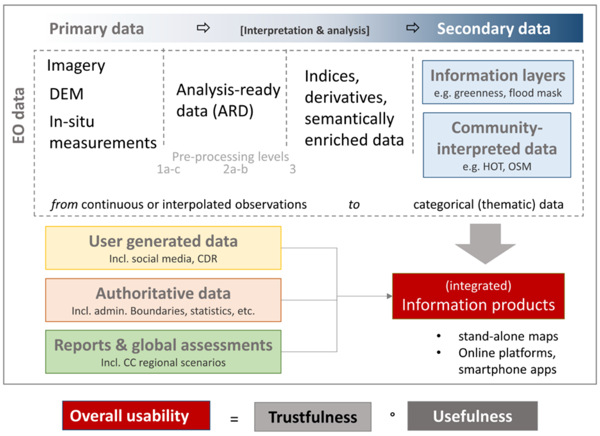

Figure 1. Data and information products. Data quality issue may influence all stages from primary to secondary, and all kinds of input data that are required for an integrated information product.

2.2 Domains and components of data quality

Different requirements for the assessment of quality and usability arise for primary vs. secondary data, for EO vs. GIS data, for collected vs. modelled data, and for authoritative data such as socio-economic data or data originating from surveys or censuses. Accordingly, there are different existing initiatives and standards ensuring data quality of different data types, e.g. the ISPRS working group on Data Quality (Batini et al., 2017) or QA4EO by CEOS; and for different aspects such as metadata standards (e.g. ISO19115), thematic standardization frameworks like the European INSPIRE initiative or technical interoperability, ensured by the Open Geospatial Consortium (OGC).

Commonly, data quality is a multidimensional concept that broadly describes the fitness for use, in particular in support to decision making as well as the conformance to set standards (Herzog et al., 2007). Next to (i) 'accuracy' as the most intuitive and well-researched quality dimension (Wang et al., 1995), there are other dimensions widely established (Mahanti, 2019), comprising (ii) completeness, (iii) consistency, (iv) timeliness, (v) uniqueness, (vi) validity; (vii) relevance, in different orders of prioritisations. The trade-off between accuracy and timeliness (Ballou and Pazer, 1995) which intuitively confronts a user with the choice between accurate but outdated data, has a specific meaning in humanitarian applications, where time-criticality over rapidly changing conditions, is key. This may also refer to information products in particular, where concepts such as "first available maps" or "fast available products" are meanwhile established service elements. In addition, there is balance between timeliness and coverage. Generally considered highly up-to-date, even standard reference data sets such as Google Maps lack behind, especially in fast onset humanitarian crisis situations, Crowd mapping (such as Humanitarian OSM or Missing Maps) provide more timely information in this specific setting while the overall (global) coverage might be limited.

Broadly abstracting from specific dimensions there are two major perspectives referring to quality structure (Meirovich, 2006), notably quality of design and quality of conformance (Heinrich et al., 2007). While the first has a more qualitative and often more subjective component addressing a potential misalignment between user requirements (of individual users or a community) and the provided data or information product, the second is more objective and characterises the correspondence of a specified data schema or measurement detail and the actual values. When referring to geospatial data (including EO data), conformance may be considered as data-inherent quality, which greatly influences the level of trustfulness: a defined maximum precision or level of detail, or degree of completeness may set the standard, the deviation of which can be quantified. One example is the top-of-atmosphere correction of remotely sensed imagery, as a key step to achieve remote sensing data calibration. Another example would be the incompleteness of OSM data in a specific urban setting. A third example is the misalignment of a set of administrative units to an existing base map. Inherent data quality is independent of the application case and can be measured according to the deviation from an (assumed) maximum value: for example, completeness in %, or offset in meters. Geospatial data usually have a well-defined or at least community-agreed inherent maximum quality level (precision, scale, etc.), such that conformance can be determined (or least estimated) quantitatively. The other aspect, , is a relative quality parameter and represents the usefulness (fitness-for-use) as a second component of the overall usability. It depends on the application case. In geospatial applications, this includes the debate on the scale of investigation, meaning that the observation scale needs to match the modelling scale (Hagenlocher et al., 2014). Example; a well-calibrated TOA corrected Sentinel-2 image might be of limited use in detecting single dwellings in a deprived urban area. It does not question its inherent quality (which might be perfect), but its appropriateness in usage in a certain application context. Usefulness, however, we hardly can measure directly, therefore we rely on expert judgement. Still, in order to compare both components on a gradual scale we need to take into consideration their different data scale levels and ideally bring them into an interval scale. As an intermediate step, we suggest a binning of 10 for each component. That means, we operationalise inherent quality by using step percentages (10%, 20%, …, 90%, 100%) taking into account its deviation from a maximum quality level (while 0% denotes maximum deviation and 100% minimum deviation), and we apply a 10-step Likert scale (1-10) for the assessment of the relative quality. Both gradual step assessments can be converted into scores. When doing a suitability assessment of several input data sets with respect to conformance (inherent quality), it obviously requires high scoring of all data in order to achieve high trustfulness in the process of data integrattion. For relative quality, this intuitively applies as well, leading to maximum usefulness at a generally high scoring level. Still, usefulness may be *fairly* high, when uniform scoring prevails on a low level. In other words, we assume that a known overall low level of relative quality is easier to handle than mixed quality levels. In geospatial applications a commensurate scale (or domain of scale) of investigation is an aspect of design quality. Low levels of relative quality might be compensated with a shift in scale of investigation rather than mixing input data with a different observation scale. An example would be an assessment of malaria breeding habitats. Ideally, the study would be conducted on high spatial detail, but due to the lack of VHR data, the scale of investigation is adapted to a coarser scale where all input data score high in usefulness.

2.3 Protocol for data quality assessment and integration

When starting the process of data integration, one often faces an unknown degree of heterogeneity. In order to minimize errors and avoid extra efforts, there is a need for a protocol, which may guide users and practitioners to perform data integration in a solid, sound and well-informed manner. This comprises the integration of EO data with VGI-generated data, survey-based field datasets and pre-existing public, institutional and administrative e.g. socio-economic data. Table 1 provides an overview on data quality aspects related evaluation scores, which enables users to assess quality and suitability of data sources. Note that for reasons of simplicity we illustrate the concept with a 4-step scoring (as opposed to 10 bins as suggested above).

The information for scoring data sources can be partly found in the metadata. Metadata record the overall quality of the data on the level of the entire record set. A simple example is cloud cover of remote sensing data, which ideally (but hardly in reality) is 0%. Automatic data quality assessment relies on data quality methods (Woodall et al., 2014) as algorithms for detecting errors and issues on validity and integrity. Global quality parameters assume the entire dataset is recorded under the same conditions and captured instantly or within a short time frame. Otherwise each record would need to be tagged with individual data quality indicators (Wang et al., 1993).

Other aspects such as completeness of features needs to be checked depending on the area of interest and information needed for a specific application case. Such an evaluation can be used for two different scenarios:

(i) Detailed assessment and comparison of different data sources describing the same aspect to find the most suitable data for a specific applications;

(ii) Evaluation of thematically different data sets that potentially should be integrated to answer specific research and practical questions at a specific location.

Table 1. Data quality protocol for integrating EO data, GI data set and spatially implicit socio-economic data

Quality dimensions [conformance]		**Evaluation score (low to high)**			
		1	***2***	***3***	***4***
① *Accuracy and Reproducibility*	EO data: thematic accuracy	<60%	60-80%	80-90%	>90%
	Declaration / reproducibility of models and methods: e.g. classification, delineation, survey design, etc.	unknown or unclear methods, no reproducibility	methodol. details unclear, low reprod.	established method, missing details	established method, fully reproducible
② *Accessibility & Exchangeability*	Open source vs. restricted data access	restricted, private use	commercial use	partly or temporary open source	open source
	Interoperability according to OGC standards; data integration as service	non-standardised format	proprietary format	exchange format	fully inter-operable and harvestable
③ *Trust of source*	Reputation of data producer[1]	unknown producer, no info available to evaluate	unknown producer, partly reviewed e.g. VGI	known, but not reviewed by community	well known, public, reviewed and certified
④ *Spatial precision*	Spatial data: accuracy of the position of features on Earth surface	no projection information, location unknown	severe shift, incomplete/ incorrect info about projection	slight shift in position, less widely used projection	accurate position, widely used projection
⑤ *Consistency / Comprehensibility*	Consistency of attributes, measure-ment levels, capturing scale, etc.	low	low-medium	medium-high	high
	Metadata documentation	documentation missing / not comprehensible	metadata not self-explanatory	partly documented, largely self-explanatory	fully documented, comprehensi ble
⑥ *Completeness*	EO data: cloud cover, degree of haziness, malfunction of sensor	>50%	25-50%	10-25%	< 10%
	Spatial features, specific attributes	<30%	30-60%	60-80%	≥80%
Quality dimensions]design]		**Evaluation score (low to high)**			
		1	***2***	***3***	***4***
⑦ *Timeliness*	Timeliness and relevance of data EO data: date of acquisition Derived data: date of acquisition of underlying data sources	≥10 years >12 months[2]	5-10 yrs 6-12 mths	2-5 yrs 1-6 mths	≤2 yrs 2-4 weeks
⑧ *Coverage / availability*	Specific area of interest: coverage of AOI	<30%	30-60%	60-80%	>80%
	General evaluation: scale of observation	local or regional	country-wide	continent-wide, various countries	global
⑨ *Spatial resolution*	EO data: spatial resolution, Derived data: resolution of underlying data sources	≥30 m	5-30 m	1-5 m	≤1 m
	Survey/census data and aggregated data: level of availability[3]	region or* country	city*	city district*	household*

[1] Organisational aspect
[2] Second row: in the context of a fast-onset crisis
[3] Here: considering an urban scale level

As an illustration for the evaluation of different data covering the same aspect, we take the example of evaluating building footprints data sets (Fig. 2 and Fig. 3) to use as input for population modelling in urban settings. We may differentiate in the context of building footprints: (i) ad-hoc generated footprints via AI-supported information extraction techniques from VHR satellite imagery; (ii) crowd-based mapping (OSM and HOT, Missing Maps and similar) and (ii) model-based or algorithm-based solutions from (semi-)public or private organisations for settlements (World Settlement Layer, Global Human Settlement layer, European Settlement Map, Building footprints, etc.).

Not surprisingly, high evaluation scores of temporal resolution and coverage are often in conflict with high scores in spatial resolution or reliability and accuracy of applied method e.g. the best overall evaluation is received by the European Settlement Map (ESM), providing consistent, complete and reliable building footprints but are only available for Europe and not regularly updated. Although OSM data has a low evaluation score in some of the aspects, we want to strengthen the importance and value of this freely available data source for research and practitioners in various field, such as the humanitarian work (Herfort et al., 2021), but also for the development and improvement of other building footprint data sets (Corbane et al., 2019).

As an example for the evaluation of thematically different data sets Fig. 4 shows heterogeneous data potentially relevant to describe city structure. Such an assessment is helpful to see if all data sources are suitable to be included in further analyses – e.g. the inclusion of the elevation model might with much coarser resolution might be questionable - and for identifying strategies that need to be implemented to harmonise and make data suitable for aggregation and/or spatial integration. This procedure can even be formalized and be used to complement statistical analyses for obtaining not only statistically sound, but also spatial and thematic sensible results.

3. DATA INTEGRATION

3.1 Data harmonisation

As data are of very different nature (scale of measurement and representation, spatio-temporal resolution, extent, etc.), data integration is more than just classical 'GIS overlay'. It requires a series of pre-processing, harmonisation techniques, strategies for disaggregation if needed, as well as (process) model integration with observations, what in total we may refer to as data assimilation (Lahoz and Schneider, 2014). The current transition to the big data paradigm also poses new challenges to existing data assimilation techniques, such as the need of integrating e.g. citizen science data to satellite-based land surface measurements on completely different scales.

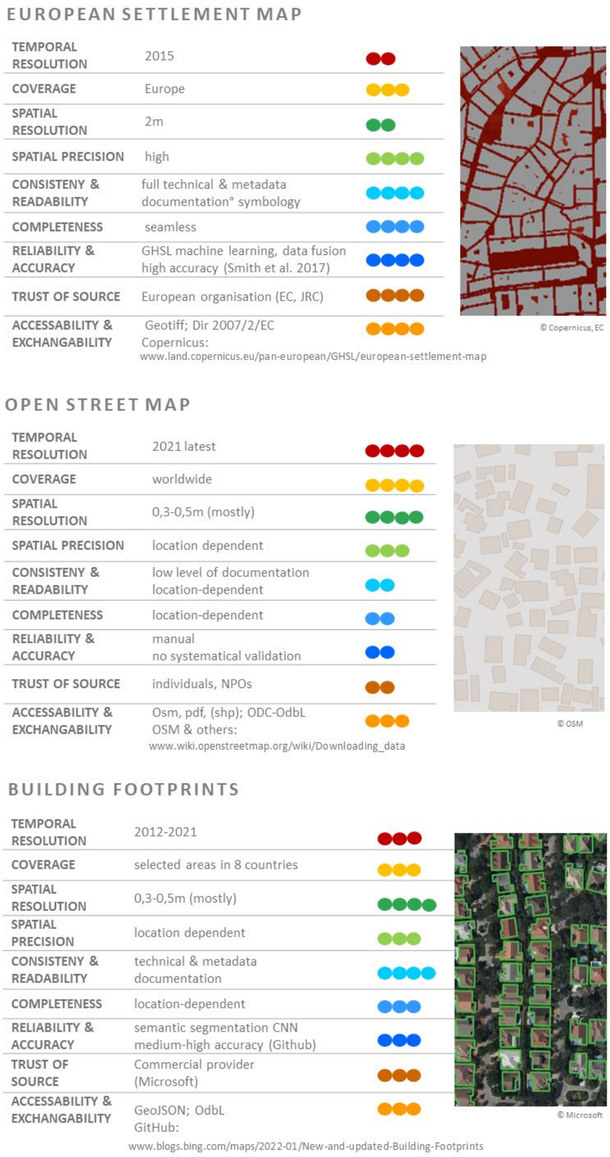

Figure 2. Detailed assessment of usability of freely available data sources of building footprints: European Settlement Map (ESM), Open Streetmap (OSM) and Building Footprints

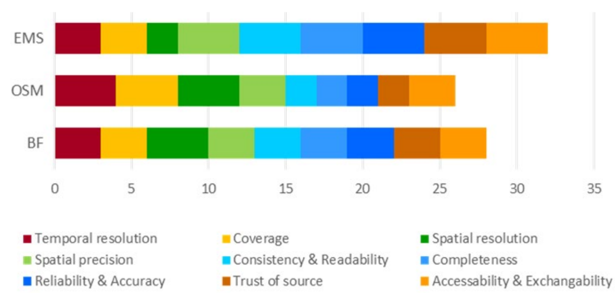

Figure 3. Comparison of usability of freely available data sources of building footprints: European Settlement Map (ESM), Open Streetmap (OSM) and Building Footprints

The first step in data assimilation routines is data harmonisation to make data comparable and ensuring successful and statistically sound data integration. Data harmonisation usually employs one or more of the following techniques, depending on the level of similarity of data:

(i) Spatial referencing / co-registration: the crucial principle that data are measured or represented in the same or re-projectable spatial reference system (e.g. UTM);

(ii) Data calibration: calibrating sensor data to ensure measurements to be comparable to each other on a defined scale (e.g. 0-100% reflectance);

(iii) Data normalisation: utilize the full value range of the measured phenomenon (e.g. 8-bit coding), may also include harmonisation of classification routines;

(iv) Data standardisation: ensure statistical data comparability through e.g. z-transformation;

(v) Data interpolation: spatial interpolation techniques to reach from a sparse sampling to area-intensive coverage.

3.2 Spatial integration

To realise data-informed development policies often also requires location-specific accurate results on a very fine resolution. One pragmatic solution is to achieve this is to use regular tessellation of the area of interest and to bring all available information on a generic discrete grid of a defined spatial reference, orientation, and spacing in multitudes (e.g. grids or hexagons). This strategy, i.e. equally spaced sampling known from image data representation (pixels) or other spatial continua such as temperature, elevation, etc. (raster cells) can be used to disaggregate other spatially extensive variables such as data reported and collected on administrative unit level, e.g. dasymetric mapping).

Thereby, data of different resolutions, both sampled and interpolated data, as well as disaggregated data, can be integrated (Hagenlocher *et al.* 2014) by which several advantages arise to enhance spatial analysis:

(i) Combination of data of different sources (including from models and observations); regular updates of data sets can be incorporated;

(ii) Harmonisation of varying resolution levels in terms of geometric properties, extent, resolution, whereby a fixed spatial reference and reporting grid allows further (re-)aggregation

(iii) Analysis of data sets in a multidimensional feature space (or data cube) by multivariate statistics and regionalisation techniques (Lang, 2018).

A second, frequently used option is the spatial aggregation by administrative boundaries, such as municipalities or an existing urban zoning. This has the advantage of using already existing reporting units (e.g. enumeration areas of socio-economic census data). On the other hand, it may distort the actual distribution of relevant information assuming a spatial homogeneity within the pre-defined units, as addressed by the modifiable areal unit problem (MAUP) (Openshaw, 1984) and its two aspects related to scaling (number of units per group) and zoning (actual grouping). This particularly applies when indicators are aggregated over admin units, which do not reflect the actual spatial distribution of the measured phenomenon.

One solution to minimize aggregation problems is spatial regionalization as discussed by the concept of geons (Lang et al., 2014). Thereby grid cells are grouped together based on a set of relevant indicators. This intuitively applies to complex settings like urban areas.

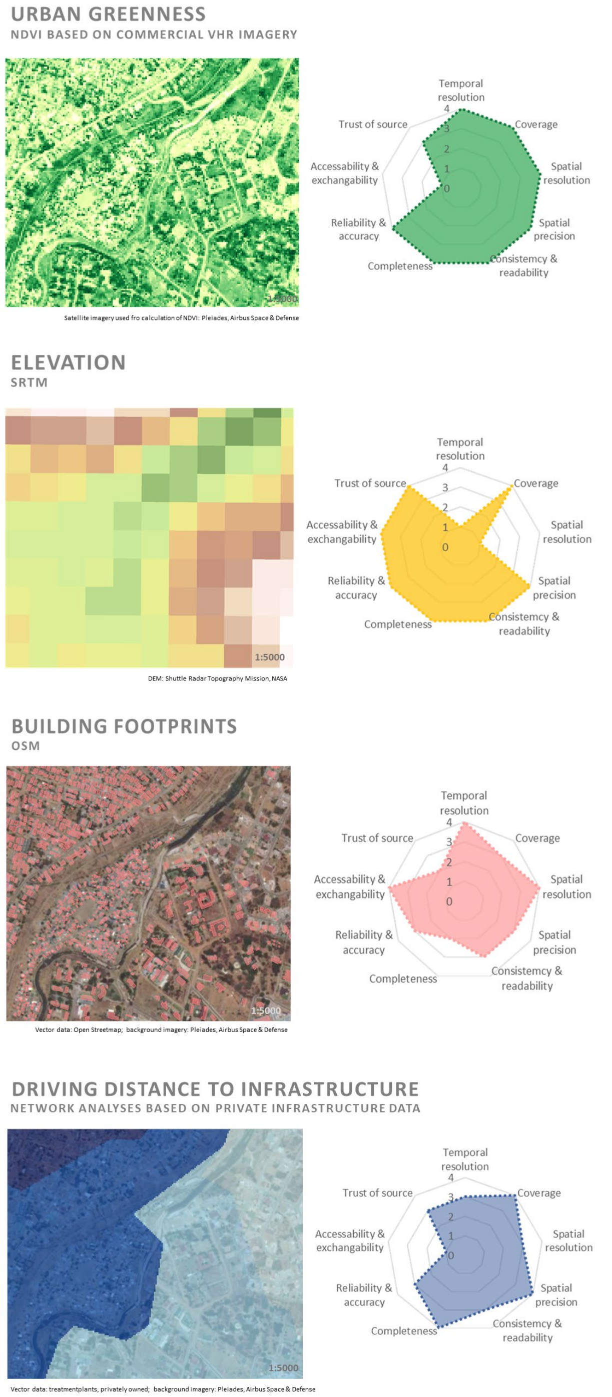

Figure 4. Heterogeneous geospatial data sets for an integrated assessment of the city structure in Lusaka, Zambia

Grouping and regionalisation realised by variance and/or spatial autocorrelation leads to statistical averaging and additional emergent spatial properties (size, form, etc.) of the generated regions. Regionalisation based on a spatial explicit multi-indicator system minimizes *a priori* spatial biases and providing spatially explicit results that are independent of pre-defined boundaries. This concept has been applied in various studies and domains (Hagenlocher et al., 2014; Lang et al., 2018; Riedler and Lang, 2018) and proves to be suitable in sustainable urban development in monitoring dynamic structure.

4. CONCLUSIONS

High-level indicators supporting the achievements of SDG 11 require, next to the methodological setup to integrate them, a solid understanding of the quality, reliability and overall usability of input data. The latter is a function of trustfulness (as the total of conformance quality aspects) and usefulness (as the the total of design quality aspects). Assuming that well-defined and (semi-)standardised data sets are increasingly available, the suggested evaluation protocol for multi-source data sets can contribute to study of complex urban features and support sustainable urban development. If the expected overall quality levels are continuously raising by various global endeavours, the potential and reliability of these datasets for integrated assessments is increasing as well. That would probably shift the attention from dealing with imperfect and partly defect data sets towards the issue of fitness-for-use. Ideally, there are multiple options offered to match various application scenarios, including detail (scale), time-criticality, and so forth. Such an application-driven view may overcome barriers in using advanced GI routines, thus bridging the gap to practitioners, such as governmental institutions, national entities or NGOs to integrate multiple geospatial information sources in their work and decision-making processes.

ACKNOWLEDGEMENTS

Research presented in this paper received funds from the Austrian Federal Ministry for Digital and Economic Affairs, the Christian Doppler Research Association, Médicins Sans Frontières (MSF, Ärzte ohne Grenzen) Austria, and the Federal Government of the Province of Salzburg (WISS 2025 initiative).
.

REFERENCES

Abascal, A., Rothwell, N., Shonowo, A., Thomson, D.R., Elias, P., Elsey, H., Yeboah, G., Kuffer, M., 2022. “Domains of deprivation framework” for mapping slums, informal settlements, and other deprived areas in LMICs to improve urban planning and policy: A scoping review. Computers, Environment and Urban Systems 93, 101770.

Ballou, D.P., Pazer, H.L., 1995. Designing information systems to optimize the accuracy-timeliness tradeoff. Information Systems Research 6, 23-51.

Corbane, C., Politis, P., Sabo, F., Kemper, T., 2019. The European Settlement Map 2019 release Application of the Symbolic Machine Learning to Copernicus VHR imagery. Publications Office.

Hagenlocher, M., Kienberger, S., Lang, S., Blaschke, T., 2014. Implications of spatial scales and reporting units for the spatial modelling of vulnerability to vector-borne diseases, in: Vogler, R., Car, A., Strobl, J., Griesebner, G. (Eds.), GI_Forum 2014. Geospatial Innovation for Society. Wichmann Verlag, Berlin, pp. 197-206.

Heinrich, B., Kaiser, M., Klier, M., 2007. How to measure data quality? – a metric based approach, in: Rivard, S., Webster, J. (Eds.), Proceedings of the 28th International Conference on Information Systems (ICIS), Montreal, Canada.

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., Zipf, A., 2021. The evolution of humanitarian mapping within the OpenStreetMap community. Scientific Reports 11, 3037.

Herzog, T.N., Scheuren, F.J., Winkler, W.E., 2007. Data quality and record linkage techniques. Springer, New York.

Hofmann, P., Taubenböck, H., Werthmann, C., 2015. Monitoring and modelling of informal settlements – a review on recent developments and challenges, Joint Urban Remote Sensing Event (JURSE 2015), Lausanne, Switzerland.

Jochem, W.C., Tatem, A.J., 2021. Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the R package foot. PLoS One 16.

Kothencz, G., Kulessa, K., Anyyeva, A., Lang, S., 2018. Urban vegetation extraction from VHR (tri-)stereo imagery - a comparative study in two central European cities. European Journal of Remote Sensing 51, 285-300.

Krauß, T., Pablo, D.A., Wendt, L., 2019. Cross-track satellite stereo for 3D modelling of urban areas,. European Journal of Remote Sensing 52, 89-98.

Kuffer, M., Pfeffer, K., Sliuzas, R., 2016. Slums from space—15 years of slum mapping using remote sensing. Remote Sensing 8.

Kuffer, M., Thomson, D.R., Boo, G., Mahabir, R., Grippa, T., Vanhuysse, S., Engstrom, R., Ndugwa, R., Makau, J., Darin, E., de Albuquerque, J.P., Kabaria, C., 2020. The role of Earth observation in an integrated deprived area mapping “system” for low-to-middle income countries. Remote Sensing 12.

Lahoz, W.A., Schneider, P., 2014. Data assimilation: making sense of Earth Observation. Frontiers in Environmental Science 2, 16.

Lang, S., 2018. Urban green valuation integrating biophysical and qualitative aspects. European Journal of Remote Sensing 51, 116-131.

Lang, S., Blaschke, T., Kothencz, G., Hölbling, D., 2018. Urban green mapping and valuation, in: Weng, Q., Quattrochi, D.A., Gamba, P. (Eds.), Urban Remote Sensing. CRC Press, Boca Raton, pp. 287-308.

Lang, S., Kienberger, S., Tiede, D., Hagenlocher, M., Pernkopf, L., 2014. Geons – domain-specific regionalization of space. Cartography and Geographic Information Science 41, 214-226.

Mahanti, R., 2019. Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. ASQ Quality Press.

Meirovich, G., 2006. Quality of design and quality of conformance: Contingency and synergistic approaches. Total Quality Management & Business Excellence 17, 205-219.

O’Connor, B., Moul, K., Pollini, B., de Lamo, X., Simonson, W., 2020. Earth observation for SDG - Compendium of Earth observation contributions to the SDG targets and indicators. European Space Agency.

Openshaw, S., 1984. The modifiable areal unit problem, Norwich.

Paganini, M., Petiteville, I., Ward, S., Dyke, G., Steventon, M., Harry, J., Kerblat, F., 2018. Satellite Earth observations in support of the sustainable development goals - The CEOS Earth Observation Handbook (special 2018 edition). European Space Agency.

Prakash, M., Ramage, S., Kavvada, A., Goodman, S., 2020. Open Earth Observations for Sustainable Urban Development. Remote Sensing 12.

Riedler, B., Lang, S., 2018. A spatially explicit patch model of habitat quality, integrating spatio-structural indicators. Ecological Indicators 94, 128-141.

Taubenböck, H., Kraff, N.J., Wurm, M., 2018. The morphology of the Arrival City - A global categorization based on literature surveys and remotely sensed data. Applied Geography 92, 150-167.

United Nations, 2021. The Sustainable Development Goals Report 2021.

Wang, R.Y., Kon, H.B., Madnick, S.E., 1993. Data quality requirements in analysis and modeling, Ninth International Conference of Data Engineering, Vienna, Austria.

Wang, R.Y., Storey, V.C., Firth, C.P., 1995. A framework for analysis of data quality research. IEEE Transactions on Knowledge and Data Engineering 7, 623-640.

Woodall, P., Oberhofer, M., Borek, A., 2014. A classification of data quality assessment and improvement methods. International Journal of Information Quality 3, 298-321.

World Bank Group, 2021. Demographic trends and urbanization, Washington.