World Settlement Footprint (WSF)

This work has been funded by the German Research Foundation (NFDI4Earth, DFG project no. 460036893, https://www.nfdi4earth.de/).

Abstract

Date of Submission 24/04/2023

Pilot title	World Settlement Footprint (WSF)
Project Duration	1 Year +
Contributors[1]	**Jan-Karl Haug**[1] (Project administration, Data curation, Writing – original draft), **Hendrik Zwenzner**[1] (Project administration), **Thomas Esch**[1] (Resources, Funding acquisition, Conceptualization, Methodology, Validation, Visualization), **Mattia Marconcini**[1] (Funding acquisition, Conceptualization, Data curation, Methodology, Software, Validation, Visualization), **Torsten Heinen**[1] (Software, Writing - review & editing), **André Twele**[1] (Data curation, Writing - review & editing), **Julian Zeidler**[1] (Software, Methodology, Data curation), **Tobias Krüger**[2] (Conceptualization, Formal Analysis, Writing - review & editing), **Sujit Sikder**[2] (Conceptualization, Data curation, Writing - review & editing), **André Hartmann**[2] (Formal Analysis, Data curation, Writing - review & editing), **Martin Schorcht**[2] (Conceptualization, Writing - review & editing), **Gotthard Meinel**[2] (Supervision, Conceptualization)
DOI	10.5281/zenodo.7897531
Version	1
Corresponding author	Jan-Karl Haug[1]

[1] German Aerospace Center (DLR), Münchener Str. 20, 82234 Wessling
[2] Leibniz Institute of Ecological Urban and Regional Development

Rapid urbanization is increasingly challenging in multiple domains including ecological, economic and social issues. In order to understand interrelationships between human habitat in relation to nature, to assess natural hazards more accurately, and to guide political and economic decisions more precisely, detailed settlement data is an important information base for further decisions and processes. This pilot project contributes novel and unique datasets from the World Settlement Footprint (WSF) suite for the NFDI4Earth community together with the IÖR-Monitor - a provider for data infrastructure and services, on land use and land cover in Germany. In addition to data itself, the data infrastructures are playing an increasingly important role to enable quick and easy access to data and a more efficient way of processing larger data. The objective of this pilot is to integrate these datasets into the NFDI4Earth infrastructure and

make them accessible to a broad user community according to the FAIR principles and common standards. Stakeholders are data and infrastructure providers, data and system integrators as well as users from different scientific domains, public authorities, and decision-makers. We expect that the NFDI4Earth will advance and highly benefit from this pilot since the offered datasets are a valuable data source for many research applications. Besides, it can be seen as a precursor for numerous other satellite data applications, services and interoperability approaches to be integrated into NFDI4Earth.

I. Introduction

More than half of the world's population is living in urban areas and the trend of rapid urbanization is ongoing. Since settlements - and urban areas in particular - represent the centers of human activity; the environmental, economic, political, societal and cultural impacts of urbanization is far-reaching. They include negative aspects like the loss of natural habitats, biodiversity and fertile soils, climate impacts, waste, pollution, natural and man-made disasters, crime, social conflicts or transportation and traffic problems, making urbanization to one of the most pressing global challenges. Accordingly, a profound understanding of the global spatial distribution and evolution of human settlements constitutes a key element in envisaging strategies to assure sustainable development of urban and rural settlements. We provide a new 10m resolution (0.32 arc sec) global map of human settlements on Earth for the years 2015 and 2019, namely the World Settlement Footprint 2015 (WSF2015) and the World Settlement Footprint 2019 (WSF2019). Additionally, the WSF Evolution and the WSF 3D are part of the product suite. The WSF Evolution is an annual time series of the World Settlement Footprint at 30m resolution from the years 1985 to 2015. The WSF 3D depicts detailed quantification of the average height, total volume, total area and the fraction of buildings at 90 m resolution at a global scale.
The datasets can be used at any scale of observation in support to all applications requiring detailed and accurate information on places with human presence (e.g., socioeconomic development, population distribution, risk assessment, etc.). In order to make the data available to a broad public, we provide a catalog service in addition to established download and visualization services. Our SpatioTemporal Asset Catalog (STAC) gives users the ability not only to search and filter data, but also to access and process it directly. This technology enables more efficient handling of data. Data transfer of large amounts of data can be prevented and storage and energy capacities can be conserved.

II. Results

Already in summer and autumn 2021 the first products of the WSF Suite could be released. The release included the WSF 2015 and the WSF 2019. At the same time, the WSF Evolution, a dataset that shows the development of settlement in a period from 1985 to 2015, was published. At the end of the pilot, the WSF 3D product was also provided. In addition to the classic map view, this data can also be visualized in a 3D view. All WSF Suite datasets show settlement and building information with global coverage. In addition to the creation and publication of the WSF products (Appendix A-1), the pilot phase also served to maintain and further develop existing data platforms. Moreover, the creation of a STAC provided a new way to describe and catalog spatiotemporal datasets. The pilot therefore gave us the opportunity to present the settlement data in a variety of ways and make the information available to a wide audience.

The different systems are also a reflection of the evolution of data access, from a service to download the data to an interoperable catalog system with which the user can not only filter data, but with the right tool, access and process the data directly without having to download it.

1) Implemented Solutions – The WSF Suite and EOC EO Products Service

The World Settlement Footprint (WSF) 2015 version 2 (WSF2015 v2) is a 10m resolution binary mask outlining the extent of human settlements globally for the year 2015. Specifically, the WSF2015 v2 is a pilot product generated by combining multiple datasets, namely:

- The WSF2015 v1 derived at 10m spatial resolution by means of 2014-2015 multitemporal Landsat-8 and Sentinel-1 imagery (of which ~217K and ~107K scenes have been processed, respectively).
- The High Resolution Settlement Layer (HRSL) generated by the Connectivity Lab team at Facebook through the employment of 2016 DigitalGlobe VHR satellite imagery and publicly released at 30m spatial resolution for 214 countries.
- The novel WSF2019 v1 derived at 10m spatial resolution by means of 2019 multitemporal Sentinel-1 and Sentinel-2 imagery (of which ~ 1.2M and ~1.8M scenes have been processed, respectively).

The WSF2015 v1 demonstrated to be highly accurate, outperforming all similar existing global layers; however, the use of Landsat imagery prevented a proper detection of very small structures, mostly due to their reduced scale. Based on an extensive qualitative assessment, wherever available the HRSL layer shows instead a systematic underestimation of larger settlements, whereas it proves particularly effective in identifying smaller clusters of buildings down to single houses, thanks to the employment of 2016 VHR imagery. The WSF2015v v2 has been then generated by: i) merging the WSF2015 v1 and HRSL (after resampling to 10m resolution and disregarding the population density information attached); and ii) masking the outcome by means of the WSF2019 product, which exhibits even higher detail and accuracy, also thanks to the use of Sentinel-2 data and the proper employment of state-of-the-art ancillary datasets (which allowed, for instance, to effectively mask out all roads globally from motorways to residential).

The World Settlement Footprint (WSF) 2019 is a 10m resolution binary mask outlining the extent of human settlements globally derived by means of 2019 multitemporal Sentinel-1 (S1) and Sentinel-2 (S2) imagery (Figure 1). Based on the hypothesis that settlements generally show a more stable behavior with respect to most land-cover classes, temporal statistics are calculated for both S1- and S2-based indices. In particular, a comprehensive analysis has been performed by exploiting a number of reference building outlines to identify the most suitable set of temporal features (ultimately including 6 from S1 and 25 from S2). Training points for the settlement and non-settlement class are then generated by thresholding specific features, which varies depending on the 30 climate types of the well-established Köppen Geiger scheme. Next, binary classification based on Random Forest is applied and, finally, a dedicated post-processing is performed where ancillary datasets are employed to further reduce omission and commission errors. Here, the whole classification process has been entirely carried out within the Google Earth Engine platform. To assess the high accuracy and reliability of the WSF2019, two independent crowd-sourcing-based validation exercises have been carried out with the support of Google and Mapswipe, respectively, where overall 1M reference labels have been collected based photointerpretation of very high-resolution optical imagery.

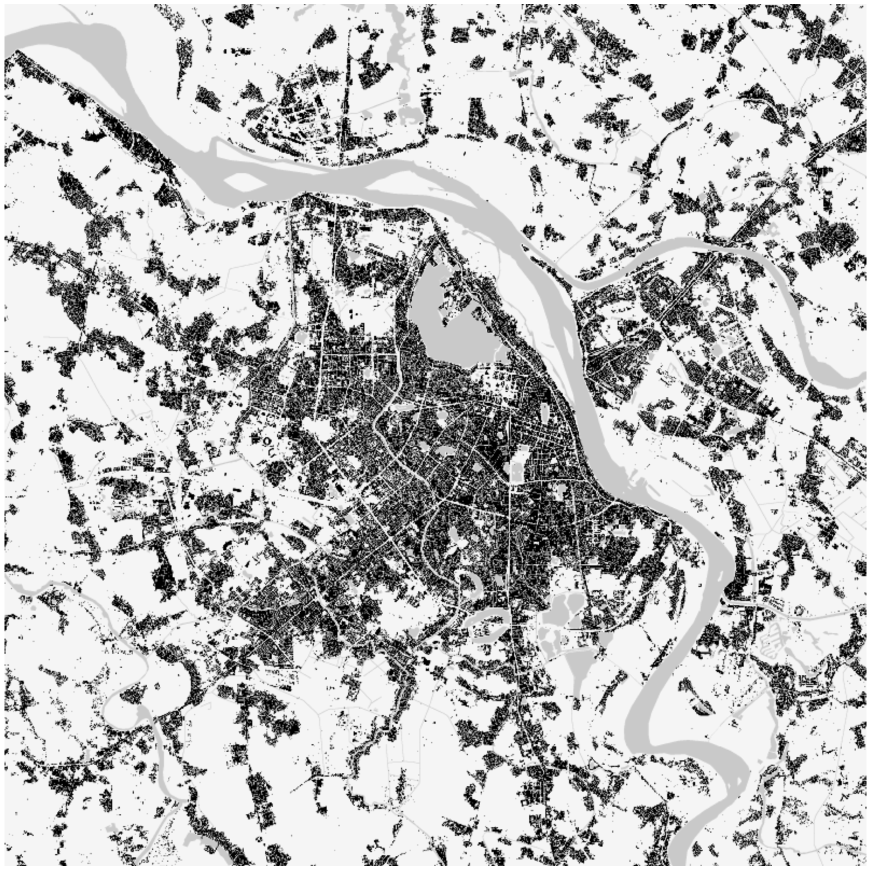

Figure 1: The World Settlement Footprint 2019 depicts the urbanization with a resolution of 10m on a global scale.

The World Settlement Footprint (WSF) Evolution is a 30m resolution dataset outlining the global settlement extent on a yearly basis from 1985 to 2015 (figure 2). Based on the assumption that settlement growth occurred over time, all pixels categorized as non-settlement in the WSF2015 (Marconcini et al., 2020) are excluded a priori from the analysis. Next, for each target year in the past, all available Landsat-5/7 scenes acquired over the given area of interest are gathered and key temporal statistics (i.e., temporal mean, minimum, maximum, etc.) are then extracted for different spectral indices. Among others, these include: the normalized difference built-up index (NDBI), normalized difference vegetation index (NDVI) and modified normalized difference water index (MNDWI). Temporal features proved generally robust if computed over at least 7 clear cloud-/cloud-shadow-free observations; accordingly, if for a given pixel in the target year this constraint is not satisfied, the time frame is enlarged backwards (at 1-year steps) as long as the condition is met. Starting backwards from the year 2015 - for which the WSF2015 is used as a reference - settlement and non-settlement training samples for the given target year t are iteratively extracted by applying morphological filtering to the settlement mask derived for the year t+1, as well as excluding potentially mislabeled samples by adaptively thresholding the temporal mean NDBI, MNDWI and NDVI. Finally, binary Random Forest classification in performed. To quantitatively assess the high accuracy and reliability of the dataset, an extensive campaign based on crowdsourcing photointerpretation of very high-resolution airborne and satellite historical imagery has been performed with the support of Google. In particular, for the years 1990, 1995, 2000, 2005, 2010 and 2015, ~200K reference cells of 30x30m size distributed over 100 sites around the world have been labelled, hence summing up to overall ~1.2M validation samples.

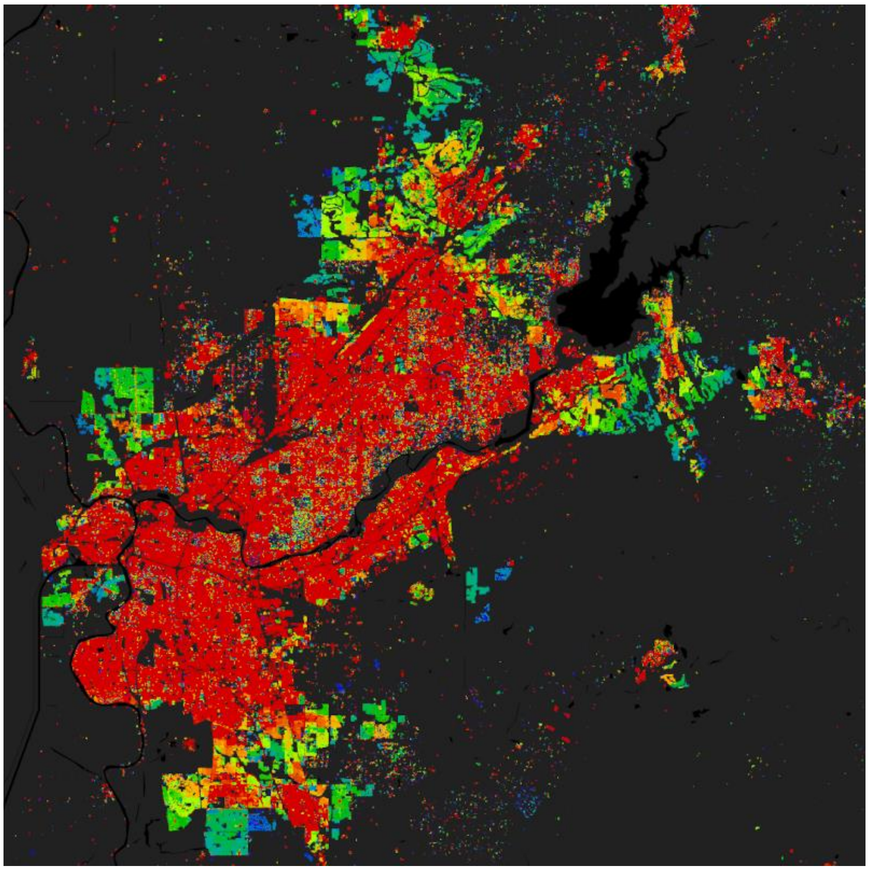

Figure 2 The WSF Evolution depicts the urbanization from 1985 (red) to 2015 (purple).

The World Settlement Footprint (WSF) 3D provides detailed quantification of the average height, total volume, total area and the fraction of buildings at 90 m resolution at a global scale (figure 3). It is generated using a modified version of the World Settlement Footprint human settlements mask derived from Sentinel-1 and Sentinel-2 satellite imagery in combination with digital elevation data and radar imagery collected by the TanDEM-X mission. The framework includes three basic workflows: i) the estimation of the mean building height based on an analysis of height differences along potential building edges, ii) the determination of building fraction and total building area within each 90 m cell, and iii) the combination of the height information and building area in order to determine the average height and total built-up volume at 90 m gridding. In addition, global height information on skyscrapers and high-rise buildings provided by the Emporis database is integrated into the processing framework, to improve the WSF 3D Building Height and subsequently the Building Volume Layer. A comprehensive validation

campaign has been performed to assess the accuracy of the dataset quantitatively by using VHR 3D building models from 19 globally distributed regions (~86,000 km²) as reference data.
The WSF 3D standard layers are provided in the format of Lempel-Ziv-Welch (LZW)-compressed GeoTiff files, with each file - or image tile - covering an area of 1 x 1 ° geographical lat/lon at a geometric resolution of 2.8 arcsec (~ 90 m at the equator). Following the system established by the TanDEM-X mission, the latitude resolution is decreased in multiple steps when moving towards the poles to compensate for the reduced circumference of the Earth. For the 3D application, the data was reprocessed. The pixel grid of the WSF3D raster was polygonized and the resulting features where attributed with the *Building_Height*, *Building_Area* and the polygon area. Based on the full resolution (zoom level 11), the polygons got dissolved and their attributes aggregated to limit the size of individual tiles in lower zoom levels.

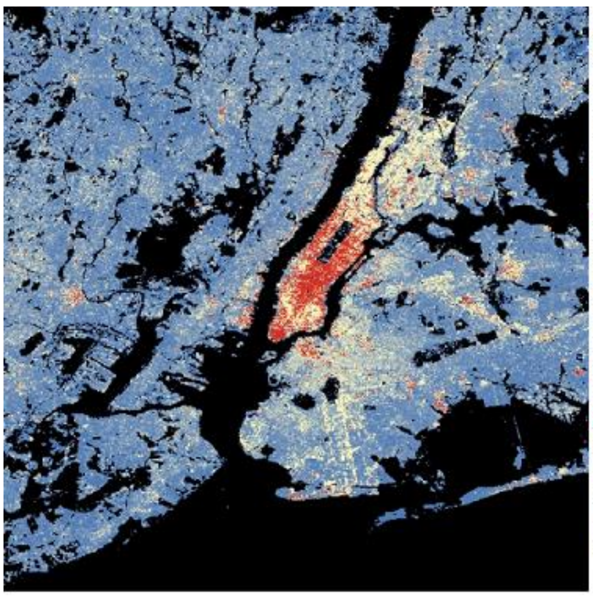
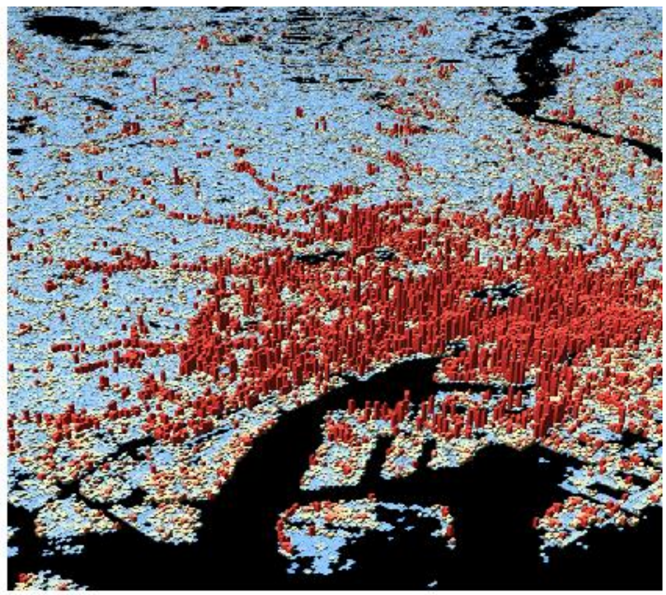

Figure 3: The WSF 3D (Building Height) for New York as GeoTIFF in 2D-View (left) and as MBTiles in 3D-View (right)

EOC EO Products Service

With large amounts of data and data sources, it is necessary to simplify access for users and data providers and to increase interoperability. The philosophy of STAC helps to make the discovery and access of data more dynamic and efficient than with previous and established data platforms, especially since many data sources are not always accessible and there are no uniform ways to search for data. In order to be able to process STAC units in a performant way, raster data are made available in Cloud-Optimized GeoTIFF (COG) format. They are structured to enable more efficient workflows in the cloud allowing the client to access and download only the parts of the file that correspond to the requested spatial area of interest and resolution.

The basis for cataloging with STAC is provided by the STAC specification. This consists of four individual specifications and provides a common structure for describing and cataloging spatiotemporal assets.

- The Catalog as a flexible JSON file of links that provides a structure to organize and browse STAC Collections and Items.
- The STAC Item representing a single spatiotemporal item as a GeoJSON feature plus datetime, properties, related assets and links.
- A STAC Collection is an extension of the STAC Catalog with additional information such as the extents, license, keywords, providers, etc. that describe STAC Items that fall within the Collection.
- The STAC API provides a RESTful service endpoint to search STAC Items with spatiotemporal and item specific properties.

Our new service, the EOC EO Products Service is already online and provides several collections, including the WSF products (Figure 4) and the IÖR settlement products for Germany (only temporarily available).

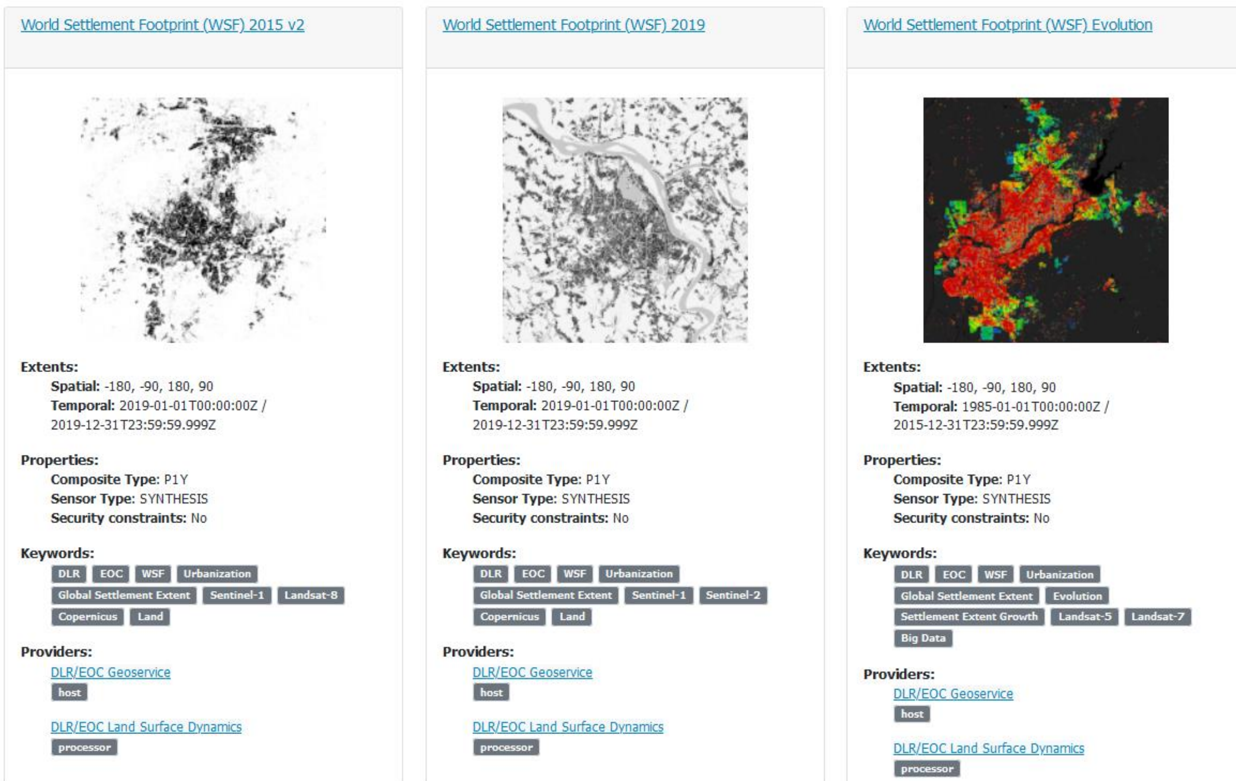

Figure 4: STAC Collection View of the WSF products (https://geoservice.dlr.de/eoc/ogc/stac/collections?f=text%2Fhtml)

For more detailed information on WSF products and STAC, several publications and links are available.

- Marconcini, M., Metz-Marconcini, A., Üreyen, S. et al. Outlining where humans live, the World Settlement Footprint 2015. Sci Data 7, 242 (2020). https://doi.org/10.1038/s41597-020-00580-5
- Mattia Marconcini, Annekatrin Metz-Marconcini, Thomas Esch and Noel Gorelick. Understanding Current Trends in Global Urbanisation - The World Settlement Footprint Suite. GI_Forum 2021, Issue 1, 33-38 (2021). https://austriaca.at/0xc1aa5576%200x003c9b4c.pdf
- Thomas Esch, Elisabeth Brzoska, Stefan Dech, Benjamin Leutner, Daniela Palacios-Lopez, Annekatrin Metz-Marconcini, Mattia Marconcini, Achim Roth, Julian Zeidler, World Settlement Footprint 3D - A first three-dimensional survey of the global building stock, Remote Sensing of Environment, Volume 270, 2022, 112877, ISSN 0034-4257, https://www.sciencedirect.com/science/article/pii/S0034425721005976
- Esch, T.; Zeidler, J.; Palacios-Lopez, D.; Marconcini, M.; Roth, A.; Mönks, M.; Leutner, B.; Brzoska, E.; Metz-Marconcini, A.; Bachofer, F.; Loekken, S.; Dech, S. Towards a Large-Scale 3D Modeling of the Built Environment-Joint Analysis of TanDEM-X, Sentinel-2 and Open Street Map Data. Remote Sens. 2020, 12, 2391. https://doi.org/10.3390/rs12152391
- https://www.dlr.de/content/de/artikel/news/2022/02/20220510_dlr-daten-zeigen-wo-menschen-durch-naturkatastrophen-bedroht-sind.html
- https://stacspec.org/en
- https://github.com/radiantearth/stac-spec/blob/master/item-spec
- https://www.cogeo.org/

2) Implemented Solutions IÖR – Building Datasets

In general, we used two types of input datasets for the processing. Building information has been derived from cadastral data excerpts on building level, whereas transportation infrastructure information has been derived from topographical data discriminating between different traffic types (road, railway, aviation).

LOD1-DE, "3D building model level of detail 1 Germany", end-of 2019)

LOD1-DE building model comprises cuboidal representations of all building objects in the Official Real Estate Cadastre Information System (ALKIS). This means cadastral building footprints are extruded by a mean or median height and covered by flat roof geometry. Furthermore, semantic attributes (though not always complete) describe building function, number of floors, roof type and metadata (Schwarz et al. 2021). The model description is defined by the Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV 2020) and Federal Agency for Cartography and Geodesy is responsible for publication (BKG 2021). They are main input data for building raster and volumetric analysis.

GA-DE, “Georeferenzierte Adressen Deutschland” (Geo-referenced postal addresses Germany, end-of 2020 release):

GA-DE are a point-based representation of building addresses of the Official Real Estate Cadastre Information System (ALKIS). Primary information is address coordinate and official address names, as described by the Federal Agency for Cartography and Geodesy (BKG 2020). They are used for main building detection and validation. For the transportation datasets, we used the following official German geo-basedata:

ATKIS Basic-DLM 2020, “ATKIS Digitales Basis Landschaftsmodell” (Authoritative topographic cartographic information system, Digital Basic Landscape Model”, end-of 2019).
ATKIS is the official object based geo-topographical model by the surveying and mapping agencies of the Federal states of Germany. The model description is defined by the Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV, 2008a, 2008b). The federal states are in charge for maintaining the information system and keeping it up to date. The object class “Actual use” of the Basic DLM provides a full and gap-free large-scale coverage of the earth surface of the administrative area of Germany. Functional land use information from ATKIS Basic DLM and address coordinates are used to validate, respectively derive, the building functions of the LOD1 objects which don’t provide building functions inherently. Traffic objects are modelled as polygons as well as polylines within the Basic DLM. These objects we used to derive the transportation footprint information.

Building footprints

Only building ground areas larger than 10 m^2 were included in the building footprint processing. Furthermore, non-residential buildings were marked with a flag. Thus, a non-residential building footprint mask could be generated additionally to the general buildings dataset. Non-residential buildings have been identified by intersecting the building footprints with ATKIS objects. The non-residential building function flag has been set for all building of which the centroids were located inside a polygon of non-residential land-use categories. Furthermore, the number of addresses per house perimeter is determined by using the GA-DE data and, based on the type of use and addressing, a decision is made as to which type of use is predominant.

The building boundaries were generated as the ground projections of the LoD1-objects. The buildings’s footprint polygons were rasterized based on the classification into residential and non-residential buildings described above.
Rasterization was done by tiles of 50x50 km tile size which have then been re-assembled to into a single grid for each of the categories general buildings and non-residential building that covers the entire federal territory respectively.

Figure 5: 1m Building Footprints (Munich, City Centre)

Transportation network footprints

In order to provide supplementary datasets for paved roads and other transportation infrastructure, which also significantly contributes to the imperviousness (soil sealing) of the land surface and is in strong functional and spatial relation to settlement structures, we also derived land covering information from linear objects of the ATKIS Basic DLM. Therefore, we buffered the transportation network according to the provided object width attributes. For the road network, the actual lane width is given by the corresponding feature attribute; within settlement areas, an allowance for sidewalks was applied (except for highways and pedestrian zones). For the railroad network, the buffer width was chosen according to the number of directional rail tracks. Roadways and airport runways also contain generalized width information through interval range attributes, which are converted into expected object widths to which buffering was applied. The processing steps are corresponded to those in detail in (Krüger et al., 2013).
The resulting polygon features are then rasterized into 1m cell size for each transportation category respecting the coordinate origin point of the LAEA projection for snapping. Transportation categories include different road dedications (Autobahn/freeway, Bundesstraße/federal roads, Staatsstraße/state roads, Kreisstraße/district roads, Gemeindestraße/municipal roads, sonstige Straße/other roads), road traffic squares, principal service roads (Hauptwirtschaftsweg), railway lines, and runways and airport aprons.

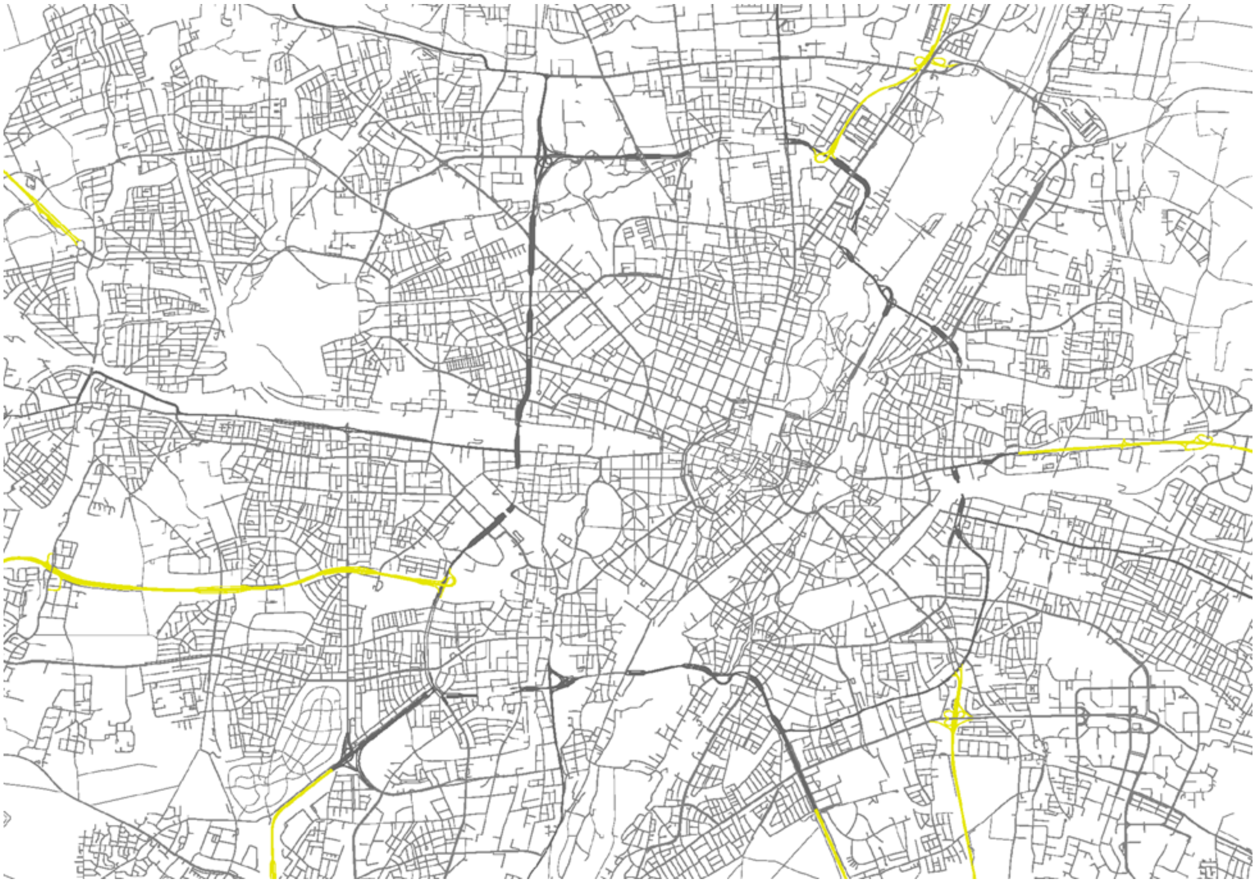

Figure 6: 1m Road Network with hierarchy, e.g. in this scene highways are colored yellow (Munich, City Centre)

Surface coverage data

The binary footprint datasets of 1m resolution have been aggregated to larger cells of 10x10 metres cell width, i.e. the sum of the included pixels can directly be interpreted as land cover percentage of each resulting pixel. Thus, each of the single datasets of the building footprints as well as the transportation footprints were processed to get surface coverage data for each category.

Development structure data

LoD1 data was used to derive development structure information for rasters of 100 m cell size. Based on the 3D-building information median building heights and the overall building volume for each cell were determined and stored as raster dataset.
The overall building volume within a cell were derived by rasterizing footprints into 1m grid cells with the building height information in metres as pixel values. Aggregating those 1m pixels into 100x100 m blocks immediately result in the volumes in cubic metres.
Median building heights within a 100m cell were assigned by evaluating the building centroids within the cell.

All above-mentioned elements of the IOER Data Suit have been stored as single raster datasets covering the full area of Germany. Additionally, each dataset also has been split into tiles of 100x100 km size fitting the INSPIRE geographical grid definitions, which makes it more convenient to integrate them into the EOC STAC. Currently, only the buildings-related datasets are intended to be integrated by the EOC STAC as accompanying data to the WSF 2019 dataset. The transportation network data, which has been developed mainly for supporting satellite imagery classificators for the discrimination of impervious surfaces into built-up and paved areas, will be separately made accessible within the framework of the IOER research data centre (IOER FDZ).

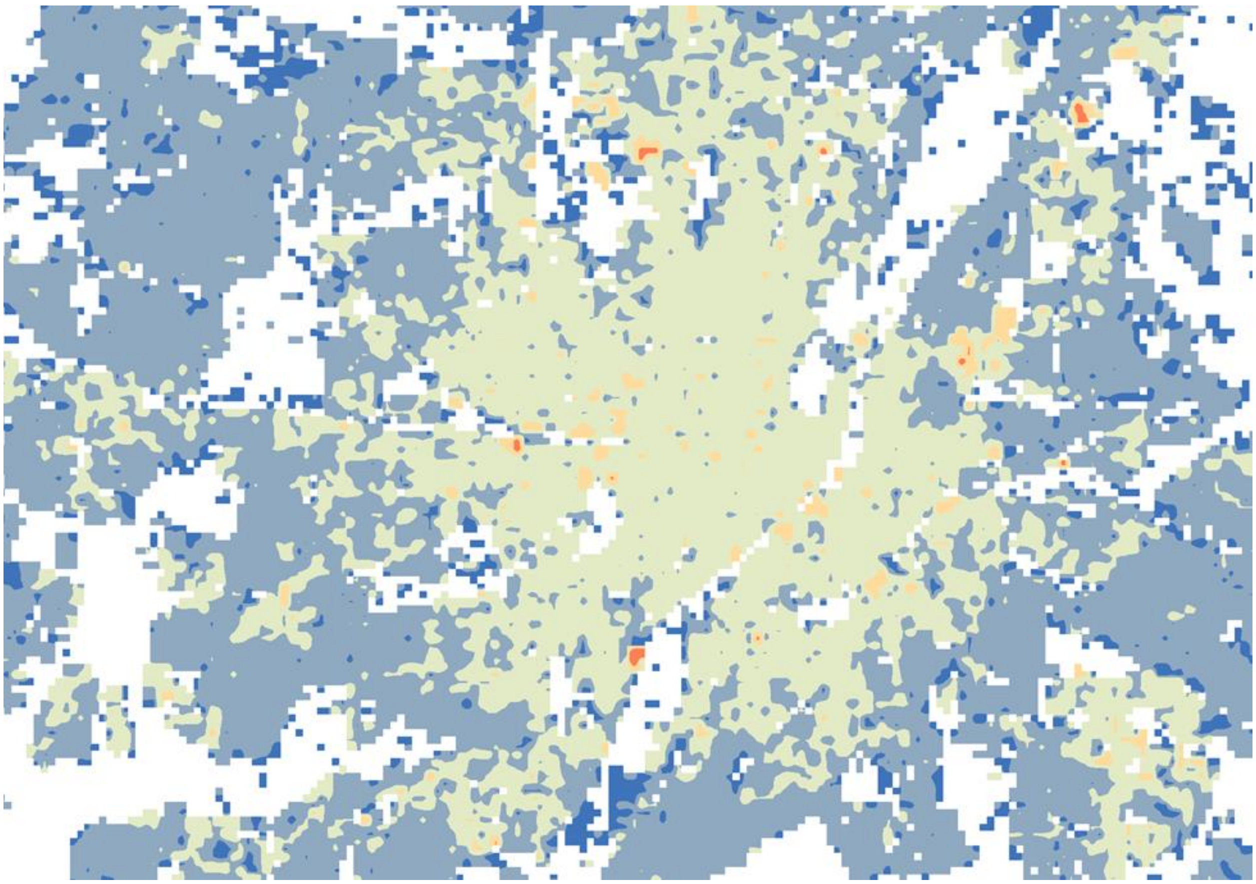

Figure 7: 100 m Median Building Height (Munich, City Centre)

For more detailed information on the input datasets (Appendix A-2, A-3) and processing building footprints, respectively, please refer to the following reference list indicating major documents and publications regarding the topic.

- AdV, 2008a. Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok). Erläuterungen zum ATKIS® Basis-DLM Version 6.0 Stand: 11.04.2008. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV).

- AdV, 2008b. ATKIS-Objektartenkatalog Basis-DLM. Version 6.0, Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok). ATKIS-Katalogwerke. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV).
- AdV, 2020. Produkt und Qualitätsstandard für 3D Gebäudemodelle. Bearbeitungsstand: 16.06.2020., https://www.adv-online.de/AdV-Produkte/Standards-und-Produktblaetter/Standards-der-Geotopographie/
- BKG, 2020. Dokumentation Georeferenzierte Adressdaten. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/ga.pdf
- BKG, 2021. Dokumentation 3D Gebäudemodell Deutschland - LoD1. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/LoD1-DE.pdf
- Krüger, T., Meinel, G., Schumacher, U., 2013. Land-use monitoring by topographic data analysis. Cartography and Geographic Information Science 40, 220–228. https://doi.org/10.1080/15230406.2013.809232
- Schwarz, S., Hartmann, A., Hecht, R., Schorcht, M., Meinel, G., Behnisch, M., 2021 Bestandsaufnahme Amtliche 3D-Gebäudemodelle im LoD1: Eine Metadatenanalyse. In: zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement 146 (2021) 03/2021, S.198-206. https://doi.org/10.12902/zfv-0352-2021
- Meinel, Gotthard, Sikder, Sujit Kumar and Krueger, Tobias. "IOER Monitor: A Spatio-Temporal Research Data Infrastructure on Settlement and Open Space Development in Germany" Jahrbücher für Nationalökonomie und Statistik, vol. 242, no. 1, 2022, pp. 159-170. https://doi.org/10.1515/jbnst-2021-0009

3) Data and Software availability

We publish the WSF products through various services for users (Appendix B). The datasets can be downloaded through our HTTP based download service both as tiles and global GeoTIFF (COG). To view the maps, we also make the data available via a Web Mapping Service (WMS) in our EOC Geoservice, as well as a Web Coverage Service (WCS) to access the raw data. For WSF 3D, another application has been prepared so that users can view the data in a 3D view. In our EOC EO Products Service, the products are also cataloged and can be accessed either through a web browser or through the RESTful STAC interface, where information about the service, the available collection and items can be queried (Appendix C). Furthermore, it is possible to limit the search geographically or temporally or to filter by product-specific attributes such as cloud coverage or values.

Another option is to use the QGIS STAC Browser plugin (Appendix D). By using the “EOC EO Product Service” URL a user can search the catalog for collections and assets and download these assets directly into QGIS.

It is also possible to integrate the catalog into a Jupyter Notebook. With certain Python packages the user is able to query the information about collections and single items, and can visualize and process the data.

The above-mentioned transportation network datasets will be made available in the repository of the IOER Research Data Centre (IOER-FDZ) as soon as it will be set up by mid-2023. Preliminary data access is available by URL https://monitor.ioer.de/fdzdata/transportation_network/.

4) Innovation and FAIRness

The amount of data and data portals is constantly increasing nowadays. This leads to a more difficult search to find the right data. In addition, access to relevant data sets is often severely limited. In order to make the search and access for EO data more efficient, more innovative methods have been developed or enhanced in the Earth observation community in recent years. To improve data management at DLR, we have used the new methods and modified our own services and would like to present these as part of the pilot. Our various services are freely accessible to users. This enables insight into the large variety of data and products from different environmental topics.

To make data *findable*, all products are marked with a unique identifier and are registered using the ISO 19115-1:2014 and the 19139:2007 Geographic information metadata standards in our Catalog Service for the Web (CSW) catalogue. The information is being harvested and is made accessible by different portals (e.g. Geoportal-DE, GEOSS, INSPIRE). Additionally, some products have a digital object identifier (DOI). Such datasets are long-term archived for at least 10 years. For the creation of a DOI, we use the functionalities of DataCite Fabrica, a web interface for registration and managing DOIs.

The EOC Geoservice with its WMS, WCS and the EOC Download Service are well-established tools to access data by visualization or download. The EOC EO Products Service adds a new technology that makes data handling more efficient. It eliminates the need for full data transfer, which is a heavy drain on computing and storage capacity, especially when downloading. With our different services we provide the framework to make data more *accessible* and *interoperable*.

Crucial to the free distribution of data is licensing. Regarding the *reusable* component, our products are licensed under the Creative Commons BY 4.0 license, that allows to freely modify and share the data without limitations, upon citing us.

III. Challenges and Gaps

We were able to achieve all set goals during the time of our pilot. In addition to the creation and publication of WSF products, we were able to provide an additional service, the STAC Catalog, where products can be easily found, explored and downloaded with the possibility of gathering one specific subset rather than the whole dataset at once. We opted for a grid of reasonable size which is a trade-off between surface covered and file size. Concerning the interoperability, we opted for a standard data type (GeoTIFF) and projection (EPSG:4326) which fit most of the purposes and can be easily transformed by the users to their preferred coordinate system.

There were no major obstacles or gaps to challenge with. So far, the Google Earth engine enabled all kind of analyses and processing. This does not hold for the 90m resolution WSF3D v1/2, which have been implemented “in house” due to the use of the WorldDEM; but the new v3 is now also generated in GEE thanks to the availability of the Copernicus GLO-30 and ALOS DEMs.

Regarding the processing of building footprints originating from authoritative datasets in order to derive the 1m high resolution building footprint dataset, we developed a workflow, which is easily reproducible in most common GIS Software allowing for future additions of data aiming in setting up time series datasets.

IV. Relevance for the community and NFDI4Earth

The WSF-Suite will be a valuable product in support to all applications and users requiring detailed and accurate information on human presence. In particular, combined either with other EO or non-EO-based datasets (e.g., related to climate, health, economy, demography, etc.), it will enable deriving indices and metrics of help not only for scientific research but even decision making. WSF products have already been used by the World Bank and other international institutions and organizations.

Apart from the various potential usages this pilot comprises several stakeholders: (1) EOC scientists, who created, collected, processed and analyzed this dataset: they benefit from the higher visibility and the wider user spectrum within the NFDI community and their feedback. (2) Data curators, infrastructure providers and system/data integrators related to the EOC Geoservice being responsible for the data accessibility of the WSF Suite: they can bring in their expertise and can benefit from tools and standards which are established in NFDI4Earth. All elements of FAIR will be addressed by this pilot. Special emphasis will be put on the quality and comprehensiveness of the metadata. Innovative interoperability approaches will be applied by our EOC EO Products Service based on the STAC. Within the framework of the NFDI4Earth pilot, it is planned to establish the IOER-Monitor as a component of the NFDI4Earth and to link its data offers in order to serve as reference information. The prime goal is to enhance the visibility and adaptability of the IOER-Monitor as a fundamental data source for spatial development issues in Germany and to achieve a quality profile as a research data infrastructure. In addition, the pilot will create an interface to the NFDI KonsortSWD by implementing a Social Spatial Science Research Infrastructure (SoRa) and work on the FAIR-compatible expansion of the IOER-Monitor.

V. Future Directions

Settlement areas play an important role in almost all disciplines of environmental research, so the expectation is that the products will be further used by a broad mass of the community. In addition, we hope to provoke some use cases of analyses or modified products using our data. For further planning, it is essential to know how valuable or useful data is and what the user's needs are.

In addition to the data, this of course also applies to the services that provide permanent access to data sets and thus ensure the long-term and sustainable usability of data. At the same time, it is important for the entire network to use the same technologies to jointly drive the further development of the Earth observation data infrastructure. This requires comprehensible documentation for the respective services and the use of the data. With the help of a repository, access to documentation and tutorials should be facilitated. These can make topics such as the connection to services (e.g. STAC, WCS, etc.) or product-specific analyses visible to all users. Furthermore, it is important to have a platform that enables an exchange with external scientists and users in order to get appropriate feedback.

The WSF suite is continuously evolving based on the recent advances on the technology, as well as the needs of the targeted communities exploiting its different datasets.
In this framework, there is a clear need for possibly maintaining all the different layers up to data and guaranteeing consistency over time. Hence, the next steps might involve the chance of systematically generating new items (where applicable) at pre-defined time steps, which shall be of particular support when investigating areas rapidly transforming as occurring in the vast majority of developing countries. On the thematic side, a key milestone for the future might be the automatic categorization of major building types (by exclusively exploiting freely accessible Sentinel imagery), which shall enable additional applications related to energy consumption, CO2 emissions and population dynamics.
In order to target the abovementioned goals, the access to the whole Sentinel-1/2 archives, as well as the possibility to perform big data processing would be required. So far shallow classification approaches have been considered. Nevertheless, the employment of deep-learning techniques is also envisaged for the future, especially when targeting the automatic building type categorization. For this purpose, suitable GPUs might be required.
The WSF suite has been taken up by a number of institutions, organizations and single users around the world as a key dataset for their everyday analyses. This is expected to further continue in the next future. The overall plan is to keep the layers updated (figure 8). The time frequency shall depend on the resources that are allocated to the activity from upcoming projects.

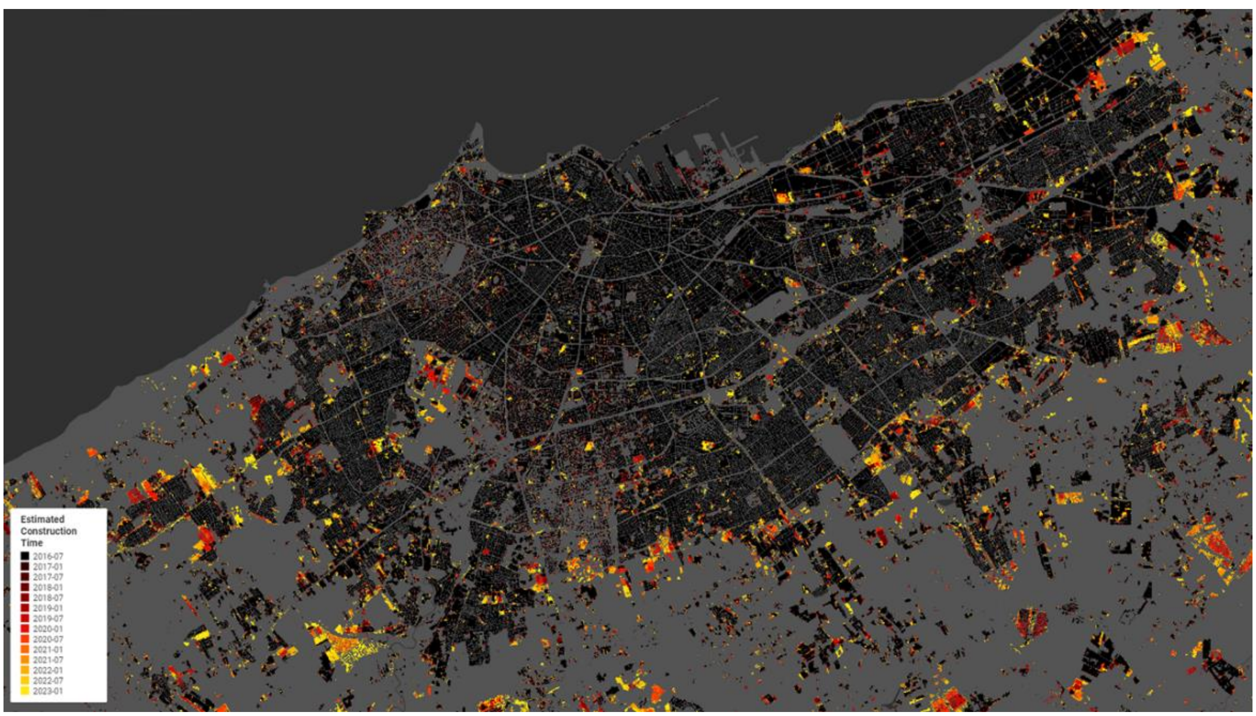

Figure 8: The image depicts settlements updated every six months (WSF Tracker - Casablanca)