NEW OPPORTUNITIES FOR URBAN LAND GOVERNANCE
BY EXPLOITING BIG DATA FROM EARTH OBSERVATION

THOMAS ESCH¹, FATHALRAHMAN ADAM¹, FELIX BACHOFER¹,², LAURE BOUDINAUD³, ANDREAS HIRNER¹, URSULA GEßNER¹, KIM KNAUER¹,⁴, ANNEKATRIN METZ-MARCONCINI¹, MATTIA MARCONCINI¹, SONER ÜREYEN¹, KARINA WINKLER¹, JULIAN ZEIDLER¹

¹German Aerospace Center (DLR), Earth Observation Center (EOC), German Remote Sensing Data Center (DFD), Land Surface Department (LAX), Oberpfaffenhofen, Germany
²Eberhard Karls Universität Tübingen, Department of Geosciences / Faculty of Science, Tübingen, Germany
³United Nations World Food Programme, Vulnerability Analysis Unit, Rome, Italy
⁴EOMAP GmbH & Co. KG, Seefeld, Germany

Thomas.Esch@dlr.de

Paper prepared for presentation at the
“2018 WORLD BANK CONFERENCE ON LAND AND POVERTY”
The World Bank - Washington DC, March 19-23, 2018

Copyright 2018 by author(s). All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Abstract

Modern Earth Observation (EO) satellite missions provide valuable opportunities to address the information needs of land governance and land use planning by delivering dedicated data on the status and spatiotemporal development of the land surface – from global down to local scale, and from urban environments to rural settings. Nowadays, satellite missions such as the US Landsat program or the European fleet of Sentinel satellites collect terabytes of high resolution imagery per day in a temporal and spatial coverage that opens up so far unprecedented possibilities for topographic mapping and environmental monitoring. But at the same time the analysis of this ‘big data from space’ requires new enabling technologies to effectively access, process, analyze and finally transform of the raw image data into ready-to-use thematic and actionable information for decision makers. Here, this contribution introduces latest developments and results of corresponding research activities that range from global mapping activities down to local applications at municipal level.

**Key Words:** earth observation, big data, enabling technologies, exploitation, actionable information
1. INTRODUCTION

Effective and efficient land governance requires detailed and up-to-date knowledge about the status and dynamics of land surface characteristics such as the land cover and land use. On this need for spatial data, the earth observation (EO) agencies respond with the continuation of successful satellite programs and new EO satellites. Covering a time span from the 1970s until today, the Landsat program provides the longest continuous record of high resolution (HR) EO satellite imagery with a current growing at a rate of ~0.5 TB per day (Baumann et al., 2016). Missions like the United States (US) Moderate resolution imaging spectroradiometer (MODIS) and Advanced very high resolution radiometer (AVHRR) operationally monitor the Earth surface with medium resolutions (~300-1000 m), which results in a data collection rate of ~30–50 GB per day. The new European Copernicus Sentinel–1/–2/–3 missions are expected to collect about 20 TB of data per day, which is an order of magnitude higher compared to the established systems (Rosengren, 2014). Hence, data of the Copernicus program can be expected to further push EO data in entering the big data era (Wagner, 2015). The above mentioned programs represent only the most prominent sources for free and open EO data. Besides those programs, an increasing number of commercial EO missions are collecting satellite images with HR to very high resolutions (VHR) below 1 m (Belward & Skøien, 2015). Planet Labs deployed more than 150 satellites with a resolution < 5 m and is able to map the entire earth at one day (www.planet.com) and other EO satellite operators like DigitalGlobe (www.digitalglobe.com) or Airbus Defense and Space (www.intelligence-airbusds.com) operate satellites with spatial resolutions of ~50 cm. Simultaneously, the availability of Unmanned Aerial Vehicles (UAV), applications and UAV-acquired data has grown exponentially (Crommelinck et al., 2016; Stöcker et al., 2017).

In view of this massive growth of available EO data, an increasing amount of highly automated data processing and analysis approaches have been developed, for instance to provide comprehensive data collections in form of large mosaics or multitemporal image composites (Roy et al., 2010; White et al., 2014; Lewis et al., 2017). The growing availability of open and free EO satellite imagery combined with enhanced spatial, spectral and temporal capabilities of running and planned satellite missions is expected to significantly improve established remote sensing applications, including land use/land cover (LULC) (Congalton et al., 2014; Fritz & See, 2008; Hansen et al., 2013; Xie et al., 2008) and urban analyses (Esch et al., 2013; Potere et al., 2009).

The objective of this publication is to present examples of thematic datasets resulting from highly automatized processing of massive data at the German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR). The paper is structured to first give examples on geospatial datasets on
global, regional and local level which are of interest for land governance and poverty applications. Next, the concept of the Urban Thematic Exploitation Platform (U-TEP) is introduced as an exemplary instrument to bridge the gap between the technology-driven EO sector and the growing need for accurate, up-to-date and actionable information as it is also required in science, planning, and policy related to land governance and poverty. Finally, conclusions are drawn with respect to the latest trends and expected future developments in terms of exploiting big data from space.

2. BIG DATA FROM SPACE – NEW GEOINFORMATION LAYERS DERIVED FROM MASS EO DATA COLLECTIONS

2.1. Global Scale Applications

2.1.1. Cloudless TimeScan Mosaics

Modern EO satellite missions such as the fleet of the European Sentinels or the United States’ Landsat missions constantly record imagery at spatial resolutions of 60-10m, adding-up to a daily volume of several terabytes of data. This massively increased global availability of medium to high resolution multispectral and SAR images provides so far unique opportunities for a detailed monitoring of the land surface in a high temporal repetition rate. However, at the same time these image collections rapidly add up to data volume that can no longer be handled with standard work stations and software solutions. Hence, DLR has developed the new processing procedure “TimeScan” which is designed to help non-expert end users to better exploit information on the type of land cover/land use from the available masses of satellite data that until now were too unwieldy for them to handle (Esch et al. 2018).

The TimeScan concept follows the idea that the individual satellite scenes are no longer transferred to the local environment of the users, as formerly, but are rather processed on large computer clusters, ideally where the mass of data were first acquired. This eliminates distribution of immense quantities of data to numerous individual users, who no longer need to have their own computer infrastructure to analyze them. Actually, now only the TimeScan end product, whose size is only a fraction of the original amount of input data, is - in case - sent to the user.

Thereby the selection of thousands of individual images covering a specific area and time interval of interest is finally condensed to one single, spatially and temporally consistent mosaic of the observed surfaces which is better suited for any further image interpretation and analytics (Fig. 1).
Figure 1: Global TimeScan-Landsat-2015 layer (top) derived from >450,000 Landsat-8 images collected in 2013-2015 and visualized as false color composite with the temporal mean of the built-up index (NDBI) in red, the vegetation index (NDVI) in green and the water index (NDWI) in blue (center); the center image images shows a zoom of the TimeScan layer for the area of Nairobi (Kenia), opposed to a representative single Landsat scene (RGB visualization) with typical situation o cloud cover and haze (bottom image).
Technically, the TimeScan processing framework includes three basic components: i) the EO2Data module for high-performance data access to EO mission archives and the ingestion of desired image collections into a processing environment, ii) the Data2TimeS module that ensures a mission specific pre-processing of the satellite imagery (e.g., in case of multispectral data cloud and shadow masking and the calculation of spectral indices for such aspects as the state of vegetation, water cover, or built-up areas, and iii) the TimeS2Stats module defining statistical ranges (such as minimal, maximal and mean value) of the various indices (calculated before by the Data2TimeS) for the selected time period of interest to generate the final TimeScan baseline product. More details related to the TimeScan methodology are provided in Esch et al. (2018).

TimeScan products have already been successfully used in various studies related to urban monitoring, land cover and land use mapping, agriculture, forestry, biodiversity, risk management and disaster prevention. In that context several million EO images (>4 PB of data) have been processed and analyzed, including global coverages of Landsat (>1,300,000 multispectral images), Sentinel-1 (>245,535 radar images), and Envisat ASAR (>25,000 radar images), as well as several regional collections of new Sentinel–2 imagery (>10,000 multispectral scenes).

2.1.2. Global Urban Footprint

The Global Urban Footprint (GUF) shows the global human settlements pattern in urban and rural environments in a so far unique spatial detail of 12m per grid cell (Esch et al., 2017). The global GUF layer, derived from >180,000 radar (SAR) satellite images (>330 TB of input data) with 3m ground resolution collected by the German satellites TerraSAR-X and TanDEM-X between 2011-2013, was officially published in November 2016. The GUF represents a binary raster dataset (built-up area, non-built-up area) that is available open and free for any scientific use and for any non-profit use in a slightly generalized version with 84m spatial resolution (Fig. 2).

To properly process and analyze the SAR imagery, the GUF production is based on a highly automated processing framework that includes a total of five modules for data management, feature extraction, unsupervised classification, mosaicking, and post-editing (Esch et al., 2017). Core of the data management component DLR’s Processing System Management (PSM) tool that allows to flexibly integrate image processing functions, to set-up work flows and to issue production requests while automatically managing the required hardware resources and workload distributions. The feature extraction module basically includes a computation of the local image heterogeneity (texture) in form of the ratio between the local standard deviation and local mean of the radar backscatter from a defined local
neighborhood. Due to the typical backscattering mechanisms occurring in built environments - the side-by-side occurrence of bright backscatter from walls and other vertical structures and dark shadow areas - the local texture shows considerably high values for settlements (Esch et al., 2011).

Figure 2: 3D-visualization of the global settlement density (left) derived from the Global Urban Footprint (GUF) data and detailed zoom of the binary GUF-2012 (top right) and experimental GUF-DenS (bottom right) for the area of Changzhou, China.

In general, the GUF allows quantitatively and qualitatively analyzing and comparing settlement patterns at various levels ranging from the local perspective to the global-scale. In this context Esch et al. (2014) introduced an advanced approach to characterize human settlement properties and patterns based on GUF-alike raster maps using techniques of spatial network analysis – e.g. in order to discriminate between urban and rural regions at continental scale. So far, more than 300 institutions from over 40 countries have requested the GUF data for a broad spectrum of applications related to global urbanization, population mapping, land governance, risk and vulnerability assessment, climate change adaptation and mitigation, or biodiversity.

Currently, DLR is working on the finalization of additional global layers providing detailed information about the local imperviousness/greenness (GUF DenS) and building volume (GUF 3D) within the
settlement area, and the spatiotemporal expansion of build-up area over the last decades (World Settlement Footprint Evolution).

2.2. Regional Scale Applications

2.2.1. Agricultural Change and Population Growth
Global population growth and changing dietary patterns lead to an increasing demand for food and to a shift towards the production of more energy- and area-demanding agricultural products. Particularly in poor regions of the world, where people heavily rely on subsistence economy, and population growth rates are high, this has led to a vast expansion of agricultural production areas during the past decades. Earth observation techniques allow quantifying and monitoring such changes in agricultural land use at national scale. The presented example (Fig. 3) illustrates the enormous conversion of land in Burkina Faso for agricultural production since the beginning of the century (Knauer et al., 2017). The maps are based on time series of satellite sensors recording in the visible and near-infrared spectrum. As these wavelengths cannot penetrate clouds, the respective images feature data gaps in case of cloud coverage. However, for an accurate mapping of agricultural areas, good data coverage for the full annual cycle is crucial as only phenological information allows for discrimination between natural vegetation (savannas) and agricultural areas. Thus, to improve data availability in cloud-affected seasons, satellite data of two optical sensors (Landsat and MODIS) have been fused (Knauer et al., 2016) and the resulting gap-free 30 m time series have been used for mapping agricultural land use. The results show an increase in rainfed agriculture from 60,441 km² (22% of Burkina Faso’s land surface) in 2001 to 114,994 km² (42% of Burkina Faso’s land surface) in 2014. But also more intensive forms of land use such as irrigated agriculture and fruit and nut plantations have increased. Even though these areas cover much smaller proportions of the country, their relative increase is considerable, with more than +340% for irrigated agriculture and almost 180% for plantations. At the same time, the population of Burkina Faso grew from approx. 12 million in 2001 to more than 17 million in 2014.

Though there are various reasons and pathways for the development of agricultural expansion and rural population growth it could be observed that these two developments go in line for multiple provinces (Fig. 4). Most of the more remote regions of Est, Sahel and Boucle de Mouhoun exhibit the highest changes in agricultural expansion and population increase. In contrast, the central, oftentimes smaller provinces of the Centre-Est, Centre-Sud, and Plateau Central only show increasing rural population numbers but no considerable agricultural expansion.
Figure 3: Changes of agricultural area in Burkina Faso between 2001 and 2014 (source: Knauer et al., 2017).

Figure 4: Scatterplots of agricultural area changes (2001-2014) versus rural population changes (2001-2014) in the provinces for of Burkina Faso (left). The colors of the points refer to the regions of Burkina Faso (right) (source: Knauer et al., 2017).
2.2.2. Monitoring of Agricultural Droughts

Droughts have serious impacts on the livelihoods of people. Particularly in the dry regions of the world, where water is a limiting factor and in countries where people strongly rely on rain-fed agriculture, droughts have serious impacts on food shortages and even humanitarian crises. As area-wide in-situ data on droughts is usually not available, earth observation provides valuable opportunities for continent-wide assessment of droughts with high spatial and temporal resolutions. In the presented example (Fig. 5, Winkler et al., 2017), agriculturally relevant droughts over Africa have been monitored between 2000–2016. Different than common drought monitoring approaches we here only focuses on droughts during growing seasons in order to assess the impact on agricultural production and food security. The approach relies on two remote sensing-based drought indices: the Standardized Precipitation Index (SPI) derived from TRMM data with 0.25° resolution and the Vegetation Condition Index (VCI) derived from MODIS data with 500 m resolution. The first assesses the meteorological aspect of droughts while the latter assesses drought effects on vegetation activity. The drought indices were combined with information on the timing of site-specific growing seasons. The results of the 17-year analysis show clear links between drought patterns of the African continent and ENSO events. To name just a few results, the years 2009 and 2011 were identified as major drought years in eastern Africa, whereas southern Africa was affected severely in 2003 and 2015/2016. Over large parts of southern Africa, droughts occurred during strong El Niño situations. A mixed drought pattern was observed in eastern Africa, where areas with two growing seasons were frequently affected by droughts during La Niña and areas with only one growing season showed droughts during the onset of El Niño.

2.3. Local Scale Applications – Socioeconomic Characterization of Urban Built-up Types

Urban areas are characterized by high-density, heterogeneous communities. In contrast to rural areas, urban inhabitants are more mobile, and their social, political, economic and institutional environment is more complex. Dynamically changing urban agglomerations in developing and emerging countries suffer from social and environmental challenges, due to a growing population and socioeconomic developments. It has been found that the traditional humanitarian sequence of assessing the needs of affected populations and then assisting these populations does not work as good as in rural areas. For instance informal settlements close to city centers or on urban fringes are often absent from maps, lack access to basic services, and are difficult to access past disasters.
Figure 5: Relative duration of drought: Percentage of growing season affected by drought events based on VCI. (source: Winkler et al., 2017).
Toward the goal of serving these communities, accurate and up-to-date maps of informal settlements are very crucial. Fig. 6 shows the result of an EO-based detection of informal settlements for the Metro Manila region based on an automated machine-learning approach (Deep Neural Network) applied to a VHR satellite image. The model was trained by using manually digitized ground truth of distinctive areas of Manila. The outputs of the model are binary maps for informal settlements like illustrated in the example in Fig. 6.

Figure 6: Informal settlements identified with a machine-learning approach from VHR satellite imagery for the area in Pasig City, Metro Manila, Philippines. The resulting binary map is overlaid on the satellite scene.

As a basis for the spatial planning of supply and disposal infrastructure, even more detailed information on qualitative and quantitative state, as well as changes, of the settlement structure is necessary. For the investigation of specific data on consumption patterns and requirements of infrastructural sectors such as energy, water, sewage and solid waste, information on the urban morphology can be used as a parameter.
to illustrate spatial patterns. Such detailed datasets of the built-up area are mostly not available in dynamically developing urban regions of the Global South.

The following example shows that these information gaps can be closed for the City of Kigali (Rwanda), by means of satellite remote sensing and household surveys. Very high-resolution multispectral Pléiades images allow identifying the number of buildings, building types and infrastructure features at a frequent basis. For the delineation of single buildings, an object-based image analysis (OBIA) approach was applied (Bachofer, 2016). Image segments are composed by spectral similar and adjacent pixels. Those segments are classified in building and other land-cover using spectral, geometrical and topologic information. The building class is then further subdivided in several distinctive building archetypes (Fig. 7) by a rule-based semi-automatized process. The overall accuracy of classification of building archetypes achieved 91%. Besides the building archetypes, urban structure types (USTs) are processed on building block level, utilizing spatial indices derived of the land-cover and building information.

Linking information gained from remote sensing with results of the household survey shows that the socio-economic status, indicated by the combined building type and USTs information, has a strong influence on the consumption patterns and requirements of the households (Bachofer et al., 2017). The quantitative and qualitative data-sets can be used to derive specific planning values and for scenarios of the future urban development.

3. AN EXAMPLE OF NEW OPPORTUNITIES ARISING FROM MODERN INFORMATION AND COMMUNICATION TECHNOLOGY – THE URBAN THEMATIC EXPLOITATION PLATFORM

The constantly progressing digitalization has led to a drastic increase of data collected on almost every aspect of our life. Thereby the combination of ‘big data’ from various sources (e.g. statistics, surveying, EO, telecommunication, social media, volunteered geographic information) with modern information and communication technologies to effectively access, manage, visualize, and jointly analyse these massive and heterogeneous data collections (e.g. via artificial intelligence/machine-learning procedures) shows great promise.

In this context the Urban Thematic Exploitation Platform (U-TEP) aims at establishing an instrument that helps bridging the gap between the technology-dominated data sector and the growing need in science, planning, and policy for accurate and up-to-date information on the status and dynamics of the built environment. Thereby it is the key objective of U-TEP to convert a diversity of raw input data sets into
“smart and fast data” (e.g. dedicated indicators) that provide concrete actionable information tailored to specific sectors and fields of application, respectively (e.g. sustainable development goals).

Core of the U-TEP system is a web-based portal that i) connects users with various data archives and repositories, ii) supports large-scale and complex data analytics and visualizations supported by high-performance computing clusters, and that iii) facilitates the sharing of data, technology, and knowledge between different user communities (see Fig. 8). Thereby the implementation follows an open source approach in a way that the interfaces between the integrated software APIs are based on open standard specifications (e.g., OCCI, OGC) and all platform components and modules are accessible via an online repository (https://github.com/urban-tep).

Figure 7: Building typology for Kigali, Rwanda. The subset shows the CBD, Nyarugenge, Kimihurura and Kacyiru.
Generally, U-TEP facilitates three basic use scenarios: first, the exploration of existing thematic products and - if required - the on-demand generation of new thematic content based on existing service applications (e.g. generating a new TimeScan product for a specific area and time of interest). Secondly, the users are provided with a development environment (virtual machine) which they can use to develop new thematic content or analysis functionalities. In a next step, the resulting new products or services can then be deployed on one of U-TEP’s high performance processing clusters in order to process large amounts of data and/or make the product(s)/service(s) available for a broader community of users. Finally, the U-TEP offers dedicated functionalities to connect and exchange with other user communities.

Figure 8: Web-portal serving as entry point to the Urban Thematic Exploitation Platform (U-TEP) and Graphical User Interface of the Visualizations and Analytics Toolbox.

A pre-operational version of the U-TEP system is already up and running since June 2017. So far more than three petabyte of data - in particular earth observation imagery - have successfully been processed with the platform. The resulting portfolio of unique and ready-to-use thematic information products and on-demand analysis and visualization services have already been used by more than 250 institutions from over 40 countries for a range of applications such as urban and environmental monitoring, sustainable development goals, disaster risk reduction, biodiversity conservation, modelling (population distribution,
vulnerability, urban metabolism), and the development of mitigation and adaptation strategies (climate change, food security, energy supply).

4. CONCLUSIONS AND OUTLOOK

In particular with respect to global and regional applications - such as given by the examples in sections 2.1 and 2.1 - the exponentially growing availability of HR imagery from modern EO satellite missions offers completely new possibilities to analyze phenomena related to land use / land cover patterns and their temporal dynamics. At the local level the recent appearance of innovative sensor platforms and concepts such as (multispectral) UAVs or High Altitude Pseudo Satellites (HAPS) adds a new dimension to existing mapping and monitoring capabilities so far addressed by means of airborne systems or commercial VHR satellites.

However, both of these trends increasingly challenge established concepts of data access, delivery, management and finally processing and analysis. At the same time it becomes more and more difficult to find efficient approaches to unlock the immense amount of information contained in such massive or new data collections for any interested non-expert user. Considering, for instance, the terabytes of data which are already collected today by spaceborne EO systems, the current concept of data distribution from the mission ground segments to the local working environment of single analysts and users who then perform individual processing steps on their personal workstations comes along with significant limitations.

One established approach to address this situation is to set-up dedicated products, services or even large-scale programs serving specific user communities – as, for instance, in form of the European Copernicus Land Monitoring Service (https://land.copernicus.eu/). However, due to the mutually “integrative” nature of such programs/products/services it turns out, in practice, that they mostly still represent a compromise for each single user or user groups with respect to the required spatial and thematic detail, timeliness, etc.. Hence, new concepts increasingly aim at bringing the processing and analysis functionalities (e.g. algorithms, toolboxes, which usually show a comparably small size in terms of data volume) required by scientific and operational communities to an infrastructure or platform where the mass volumes of EO data and auxiliary information from prominent EO missions are available in combination with high-performance processing, analytics and visualization capabilities. Well known examples for such platforms and services are the Google Earth Engine (GEE) (Gorelick et al., 2017) and the Amazon Web Services (AWS) (Amazon Web Services, 2017). Such services can be used, among other analyses, for the processing of highly automated, standardized and scientifically evaluated thematic datasets based on big earth data (Hansen et al., 2013; Pekel et al., 2016; Trianni et al., 2014).
However, the handling of these platforms and solutions still requires certain programming and other technical skills. Hence, users are ideally provided with all-in-one solutions that include adequate processing capacity, full access to multiple data archives/collections, “ready-to-use” thematic datasets, and the possibility of generating, analyzing and visualizing new datasets on-demand for the specific topic, time or area of interest. Thereby the underlying functionalities can be accessed and executed by EO experts, as well as by user of various other fields. The idea of such an end-to-end environment is addressed by the concept of the Thematic Exploitation Platform (https://tep.eo.esa.int/) - such as the U-TEP briefly introduced in section 3. These platforms pursue the objective of improving the decision making related to the sustainable management of human settlements and processes and thereby help to develop a more comprehensive and sound knowledge base for a specific phenomenon, topic or community.

ACKNOWLEDGEMENTS

The authors would like to thank the Bavarian Ministry of Economic Affairs and Media, Energy and Technology for funding the OPUS-GMES project (Concept of an Operational Platform for the Provision and Processing of Sentinel Data in Support of Copernicus Geo-Information Services) which provided the basis for the technical developments of the TimeScan framework. The authors also thank the European Space Agency (ESA) for funding the project “Urban Thematic Exploitation Platform – TEP Urban” (ESRIN/Contract No. 4000113707/15/I-NB).

REFERENCES


