Technische Universität München

The Data Integration Challenge in Smart City Projects

Updated Version, August 2020

A Whitepaper written by

Prof. Dr. Thomas H. Kolbe
Mandana Moshrefzadeh, M. Sc.
Kanishk Chaturvedi, M. Sc.
Dr. Andreas Donaubauer

Chair of Geoinformatics
Dept. of Civil, Geo and Environmental Engineering, and
Dept. of Aerospace and Geodesy
Technical University of Munich
Arcisstraße 21
80333 Munich
Germany

This document has been created in the context of the projects "Smart Sustainable Districts" and "Smart District Data Infrastructure" funded by the Knowledge and Innovation Community on Climate Change and Mitigation (Climate-KIC) of the European Institute of Innovation and Technology (EIT). It explains key requirements and considerations that any kind of infrastructure for urban data integration (like Smart City platforms) has to deal with. Within the SSD project, the Chair of Geoinformatics at TU Munich has developed an architecture called "Smart District Data Infrastructure" (SDDI) together with the SSD project partners. The SDDI has been implemented for a number of districts of larger European cities including London, Paris, Berlin, and Utrecht.

See www.climate-kic.org and https://www.lrg.tum.de/en/gis/projects/smart-district-data-infrastructure/

Preface on Data Integration for Smart Districts and Smart Cities

The word "Smart" in the Smart Sustainable Districts (SSD) Programme of Climate-KIC of the European Institute for Innovation and Technology (EIT) refers to innovative and clever ways to achieve factor 4 improvements[1] in urban districts concerning the reduction of resource and energy usage, reduction of waste, reduction of emissions, improvements of well-being, mobility, and in general concerning efficiency. In most cases, such improvements can only be achieved by concerted actions touching on different thematic fields and involving different stakeholders. In order to plan and perform such actions, well informed decisions have to be made taking into account multiple information sources from different stakeholders and their data resources. Also, the monitoring and rating of specific district performances via dashboard systems and key performance indicators require the (online) combination of data resources as well as sensor systems from different sources. Hence, providing answers to the question how the diverse data from the multiple sources within a district can be made accessible, protected, brought together, and analysed for mutual benefit is key for the success of the SSD Smart District Data Infrastructure (SDDI) approach.

Many different stakeholders, many different applications

The general situation of districts is that there are a number of stakeholders – such as citizens, public administration, transportation companies, energy providers, construction firms, housing companies – who all have their data, but in isolated form. Often the data is not accessible and the structure and semantics (i.e. the meaning) are not described or standardized. In many cases, individual stakeholders already provide or use specific data and sensor platforms which serve the needs of the respective stakeholder very well, but which are not interoperable with the platforms of other stakeholders. Different data formats, web service interfaces, and data security strategies hinder the integration of the information for joint analysis, planning, monitoring, and the preparation of well-informed decisions. While for an individual district the data and web service interfaces can be programmed for a specific application (like a dashboard application), these interfaces are neither transferable to other districts nor to other applications within the same district. Each application has to deal separately with the very specific information system scenery of the district making the development of applications a) very expensive, and b) non-transferable or non-usable with other districts that possibly have the need for a similar solution.

Data integration requires interoperability between distributed resources

In order to plan and to perform urban transformations, many stakeholders are affected and have to work together. This requires the combination and mutual usage of information resources from different partners in the development. Since each individual urban transformation project typically involves different stakeholders, it does not make sense to try to collect all available data resources from all stakeholders in advance within a central city data repository. This is also inhibited by data privacy concerns and by

---

[1] von Weizsäcker et al. 1997: Factor Four: Doubling Wealth, Halving Resource Use - A Report to the Club of Rome. Earthscan Ltd.

competition between several stakeholders. This means that the original data resources remain with their producers / owners. Data will only be brought together and integrated for specific applications or projects, in which the participating stakeholders have explicitly agreed to this. These requirements imply that a technical framework for data integration must take into account an architecture of distributed resources and functionalities which can be queried and linked to each other on demand. However, when such agreements have been made, the technical integration should be done in a quick and easy way. This can be achieved by the usage of international standards for the modelling, description, and representation of the data resources as well as for interfacing the distributed components that give access to data, visualizations, and analysis tools. When systems can interact on the level of common interfaces and data exchange formats they have achieved what is called *syntactic interoperability*.

But data integration is not only about providing access and interfaces to distributed data resources. This can be seen just as the lowest technical barrier to overcome. Once datasets have been retrieved (this may include the observation data from sensors), the receiver must be able to fully understand the meaning of each data item in order to be able to make sense of the data and to use them properly in calculations and analyses. This requires that not only the data exchange formats have to be standardized, but also the meaning of the information pieces encoded in the format must be documented and the documentation being made accessible to actual and possible users. This aspect is called *semantic interoperability*; it ensures that different users and systems understand data in the same way. Semantic interoperability can also be achieved by using international standards for domain specific data models and/or ontologies.

Most data resources are related to objects of the physical environment

When looking at multiple urban transformation projects and respective smart city applications at the same time, one can see that many applications require data about the same entities. This opens up opportunities for synergies, because specific data entities may be grouped and provided in a unified way avoiding redundancy and repetitive data acquisition. In the course of the SSD project the data requirements (and offerings) of the four “deep dive districts” in London, Berlin, Utrecht, and Paris were carefully analysed. It was found that the vast majority of required and provided data was about objects of the physical environment like buildings, streets, pathways, vegetation etc. In many cases data was directly related to some characteristics of these objects (location, size, extent, volume, ground footprint) or to some parts of these objects (e.g. in the case of a building the areas of a building façade, a roof surface, or the total living space). In other cases data was just related to objects of the physical environment, for example, the heating energy demand, the assessed real estate value, as well as the CO_2 emissions of a building. If there is a virtual model of the district (or an entire city) that represents all relevant physical objects by corresponding data objects in a sufficient degree of thematic and geometric detail, many (and sometimes all) required data items for different applications can directly be derived from such a model and be used in calculations (like the derivation of urban indicator values). Furthermore, if different stakeholders relate data items or sensors to the same object in the virtual district model (VDM), these data items become directly comparable – because they are referring to the same denominator (e.g. the same building object of the virtual district model) – and, thus, can be easily combined and jointly evaluated.

Virtual 3D models of cities, districts, and individual buildings

Virtual models of the physical environment have been investigated and developed in the fields of geospatial information science (GIS), computer graphics (virtual and augmented reality, VR and AR), computer games and training simulations as well as in the field of civil engineering and architecture. Virtual reality models as used in computer graphics and computer games are not suitable to support the alignment of information, because they are structured according to the efficient and realistic rendering of 3D scenes. In most cases only 3D geometry and appearance properties are represented, but no classification of objects into thematic classes like buildings, trees, roads etc. is being done.

In the field of GIS, semantic 3D city models have been developed, which are structured according to thematic and logical considerations instead. The international standard CityGML issued by the Open Geospatial Consortium (OGC) defines a semantic data model for the representation of the most relevant urban objects like buildings, trees, streets, vegetation, water bodies, bridges, tunnels, city furniture etc. By defining the meaning of all object classes, their attributes, and their interrelationships CityGML achieves semantic interoperability between users, data providers, and applications. CityGML also defines an exchange format and can be queried and modified using a standardised web service interface (called Web Feature Service, WFS), which establishes syntactic interoperability. More and more cities today create and maintain 3D city models according to the CityGML standard. Among others, the following cities already have a CityGML-based 3D model: Berlin, Munich, Frankfurt, Hamburg, (in fact most larger German cities), Singapore, Zurich, Vienna, Geneva, Brussels, Rotterdam, New York, Paris, Lyon, Helsinki, and many more.

Semantic 3D models play also an increasing role in civil engineering and architecture. Building Information Modelling (BIM) is a method using digital models to support the planning, construction, operation, as well as retrofitting and abolishment of buildings and other man-made objects. With the so-called Industry Foundation Classes (IFC) the organization BuildingSmart International (bSI) issued an international standard for the representation of BIM models and their data exchange. A number of countries like USA, UK, Germany, Norway and others have put regulations into force that require that new buildings or infrastructure objects must be planned and constructed following the BIM method in the next years. This means, that in the near future for all newly planned and constructed buildings, very detailed semantic 3D models will become available with standardized information contents. All kind of information related to these buildings can then be related to the BIM models facilitating data integration on a semantic level.

Virtual models of the physical reality also play a key role when it comes to the estimation and assessment of the impact of planned urban transformations on costs, quality of life, environmental and energetic parameters, but also on mobility. Since many of these parameters and indicators can be estimated by simulation tools entirely based on enriched virtual district / city models, such estimations can also be performed for 3D models that have been modified according to some planned urban transformation, representing “what-if” scenarios. For example, the impact of the retrofitting of 30% of all residential buildings in a district and the installation of photovoltaic solar panels on 25% of all suitable roofs and facades on energy production, consumption, and potential reductions of CO_2 emissions can be computed and put to relation with the costs and required materials. Also the impact of a change of traffic flows of some streets onto the energy demands, traffic noise, and air pollution, can be estimated and assessed.

Core services, which have to be operated centrally

While most data resources and analytical tools will be offered within a distributed network of components by the different partners, some core services and datasets will have to be managed and provided centrally by one of the stakeholders. The latter could be the municipal administration or a commercial company. They run a registry, in which all partners and interested parties can register and offer their data resources, applications, tools, sensors, etc. The registry can be queried and browsed by users like a catalogue to find appropriate offerings for their specific interests, needs, and applications. Also the virtual district / city model should be maintained and provided in a central way (e.g. by the cadaster or GIS department of a city), because it is being used and referenced by most applications and linked from other data resources.

Open Source or commercial software is not decisive – Open Standards are

The challenge of building up an infrastructure for urban data integration is not decided on the question whether to employ Open Source or commercial software. The key point is to require all types of software to implement and follow Open Standards. This facilitates full competition between software products no matter whether they are being manufactured and licensed by companies or created and distributed freely by some developer communities as Open Source.

There are many arguments in favour of using Open Source software: e.g. no licensing fees, unrestricted access to the source code, and often big international developer communities. Open Source implementations are also helpful to have a low entry barrier and to motivate people within the academic world as well as individual enthusiasts to develop services and contribute with data resources. However, there are also good reasons to employ commercial software. For example, certain software tools like complex simulation systems have been created based on deep special knowledge of some companies. The functionalities provided as well as the efficient way how these functionalities are implemented often are making up the competitive advantage and, hence, the economic basis for the software manufacturer. For commercial software support is guaranteed – at least for a specific maintenance time period. In many cases, commercial software is addressing more the aspects of user-friendliness and robustness (especially regarding user interfaces) and scalability than Open Source. Anyway, there exist mixed models where companies are offering user support, tailoring, branding, roll out support, and maintenance for Open Source software. As a conclusion, a data infrastructure for Smart Districts and Smart Cities should allow the usage of both commercial as well as Open Source software. The key in building up the data infrastructure is to ensure interoperability and to avoid vendor lock-ins, which is achieved by using Open Standards for data models and all relevant web services.

Open APIs do not solve the problem as long as they are not standardized

Many commercial companies offer access to the functionalities of their software systems by open application program interfaces (Open APIs). By providing openly documented APIs the companies allow users and programmers to operate, control, access, and integrate the respective software systems from / within their own applications. While the provision of Open APIs is clearly an added value making the usage of specific software systems more attractive, it does not solve the problem of data integration in Smart Districts and Smart Cities as long as the API and the employed data models and exchange formats are not standardized by a recognized standardization organisation.

The program code of an application using or embedding functionalities of a third party software system has to rely on the API and, thus, becomes dependent on it and the company who is defining it, which can cause vendor lock-ins. It is important to remember that Open APIs can be changed, modified, and being replaced by the owning company anytime and without consultation of customers and users. In fact, this happens regularly, for example, with the Open APIs of software solutions of big companies like Google and Microsoft. If one wants to build an application linking e.g. six different platforms or systems, the application developer will have to monitor and adapt to changes of six APIs, which basically requires frequent updating of that application just due to changes in the interfaces to third party systems – possibly every couple of weeks. This level of maintenance typically can only be sustained by application or solution providers who can dedicate a part of their development team to such tasks. This is hardly manageable for smaller companies and IT departments of public institutions.

Sometimes companies with Open, but proprietary APIs are pointing out that they are using standards, like JSON or XML for data encoding or REST-based web service interfaces to connect to their platforms. This, however, does not mean much. Practically all software platforms – including the ones using standardized APIs, data models, and exchange formats – are using these basic technological Web standards. Still all proprietary Open APIs of the different companies are incompatible with each other due to different protocols, different structures and meanings of the data items, and different access control and user management mechanisms.

One solution to attach proprietary systems offering Open APIs is to build translators or facades that offer standardized APIs, data models, and exchange formats on the one side, and which connect to the Open, but proprietary API of the system on the other side. Applications can then use standardized APIs to work with proprietary platforms and systems. In the Smart District Data Infrastructure (SDDI) that was developed in the Smart Sustainable District (SSD) project this approach is applied in the field of Internet of Things (IoT) for the multitude of Open APIs for sensors and their observations. Many big players like Google, Amazon, Microsoft, and IBM offer Cloud-based storage, management, and analyses of sensor data, but each of them is offering their own API. In addition, many smaller companies and initiatives in the IoT domain like Thingspeak and OpenSensors offer similar functionalities, again with their own APIs and different levels of analysis capabilities. In the SSD project, a light-weight Open Source software called "InterSensor Service" was developed and implemented, which offers the standardized APIs defined in Open Geospatial Consortium's "Sensor Web Enablement" (OGC SWE) to application developers and connects to many different proprietary APIs like IBM WUnderground, Thingspeak, Twitter, but also simple relational databases and even spreadsheets (making tabulated data stored in local files accessible as virtual sensor observations). Since the InterSensor Service is Open Source[2], further interfaces to proprietary Open APIs can be added easily (requiring typically 1-2 days of implementation).

Open Data, private data, and data security

Any framework for comprehensive data integration needs to be able to support both Open Data as well as restricted access to private data. Not all data resources can be made public and accessible as Open Data for many reasons including security, competition, data privacy, and business considerations. For example, smart meter readings of buildings or the operational status of some machines, transportation systems etc. have to be protected against unauthorized access. Furthermore, data which is directly related to individuals are not allowed to be made openly accessible without their consent due to data privacy regulations in most countries.

Nevertheless, there are reasons for two or more stakeholders having restricted data to cooperate and exchange such data, when they decide on creating specific applications and use cases requiring mutual access to restricted resources. One example could be the access of security forces like the fire department or police to operational data of transportation providers or energy companies in cases of emergency. Another example could be the cooperation of energy companies with mobility providers in order to build-up and operate an electrical vehicle infrastructure. Enabling data security is key to guarantee data privacy, both for individuals and entire organizations.

In order to facilitate the provision of commercial added value services, besides data security also accounting must be supported. This is required, for example, to charge clients according to their usage of some services as well as to document usage statistics on the provider side.

Data security for private data must be ensured by supporting proper web standards for encryption, access control, authorization, and accounting. A specific challenge that has to be tackled in the context of federated infrastructures like the SDDI is to provide single sign-on functionalities over the distributed resources operated and provided by the different stakeholders and providers of the SDDI. Multiple identity providers (IDPs) must be allowed and usable. Some resources may require that a user just has an account with some social network like Facebook, LinkedIn or Google. Other resources shall only accept requests from people and services being authenticated by some company or organization. In order to facilitate all of the above aspects, the SDDI realizes a so-called Access Management Federation (AMF) with user Single-Sign-On support based on the following well-known and widely used international IT standards: IETF RFC standards HTTPS and OAuth2, OASIS standards XACML (and the OGC standard GeoXACML) and SAML2.

[2] https://www.intersensorservice.org

Recent developments like the Blockchain technology can help in the future in managing and securing data transactions. The digital models of the environment (virtual 3D city models, BIM datasets) could be operated following the idea of distributed ledgers, where distributed copies of the same digital world representation are spread to multiple stakeholders, and using Blockchain technology to keep track of changes and to facilitate data synchronisation.

Last, but not least: Business opportunities

The Smart City market is growing and is very attractive for technical innovations and innovative businesses. The crucial point here is, although even the districts in a city are diverse and reveal different characteristics and challenges, many data integration aspects are common to different districts, cities and countries. In general, the building blocks of a Smart District Data Infrastructure remain the same, while the choice of specific software products and their composition to perform specific tasks will be locally dependent. In order to have a replicable and scalable solution, employing open and well developed standards, which are already used and tested in many places and use cases, is of high importance. Due to the fact that software tools are not just using the same application program interfaces (APIs), but they are also referring to the same object definitions using standards like CityGML, IFC, OGC Sensor Web Enablement (semantic interoperability, see above), such tools become directly applicable within and, hence, transferable to other districts and cities. This makes it beneficial for software developers to follow the standards, because there is a huge number of possible customers. Developers of applications like e.g. a city dashboard can sell their software in principle to all other districts and cities which are running the SDDI. Developers of simulation tools like energy demand estimation, solar potential analyses, traffic simulation, air pollution and noise dispersion, flood risk assessments, and asset management applications can link their tools to such infrastructure reducing the local costs for the individual district / city for the adoption of these tools.

In the context of SDDI we see markets for data, applications, analytical tools, sensor services, solution providers, integrators, consulting, education and training. Hence, there are business opportunities for data creators, providers, brokers, and refiners; developers and manufacturers of applications and tools. This includes start-ups, medium sized companies as well as big players. The many relationships between the businesses, including even small ones, will foster and ensure the stability of that framework. Stakeholders from public administration, private companies, and agencies coming from different areas like urban planning, mobility, security and disaster management, architecture, civil engineering, project developers are on the one side customers and on the other side can also become a supplier, if they create solutions for specific tasks using the SDDI framework, which they can in turn then offer to other districts and cities that also employ the SDDI. Citizens can also be users of and contributors to the SDDI. For example, citizen science initiatives could participate in providing access to crowd sourced data or sensors and at the same time exploiting resources provided by other stakeholders. Besides free usage of Open Data there is also the possibility to establish data cooperatives, giving mutual access to their resources. Such cooperatives could request regular fees or usage fees for their resources from their members.

Often there is the situation that stakeholders are not sure whether to follow certain standards. Data providers are hesitant, because of a lack of a broad implementation base of some data standards in software systems. Software manufacturers and developers are hesitant to implement the standards, because of a lack of datasets that are represented according to the standards. This chicken-and-egg problem can impede business and the growth of a market. In our developed framework of the Smart District Data Infrastructure the situation is different. The usage of Open Standards from the field of Spatial Data Infrastructures (SDI) already gives access to a wide range of existing data resources, operated by municipalities and organizations like mapping agencies and environmental institutions on the state and national level. Also global remote sensing and earth observation data from commercial providers and public agencies like NASA and ESA are immediately usable, because they offer access to their services and products using the same Open Standards. Also resource catalogs exist on regional, national, and European level which can directly be linked with the catalog service for a city or city district. Using the Open Standards from the field of SDIs also means that there already exist numerous commercial and Open Source implementations of the employed Web Services for distributed access to raster based data (maps, aerial and satellite images, weather data), geospatial objects (like 3D city models, cadastre and topography objects; in principle, arbitrary object structured datasets), catalog services, sensor services (air quality sensors, flood water level gauges, traffic density sensors in the streets etc.) and many more. The groups and companies who are providing these software tools are candidates to directly participate and compete in the market and enter the business field of Smart Cities.

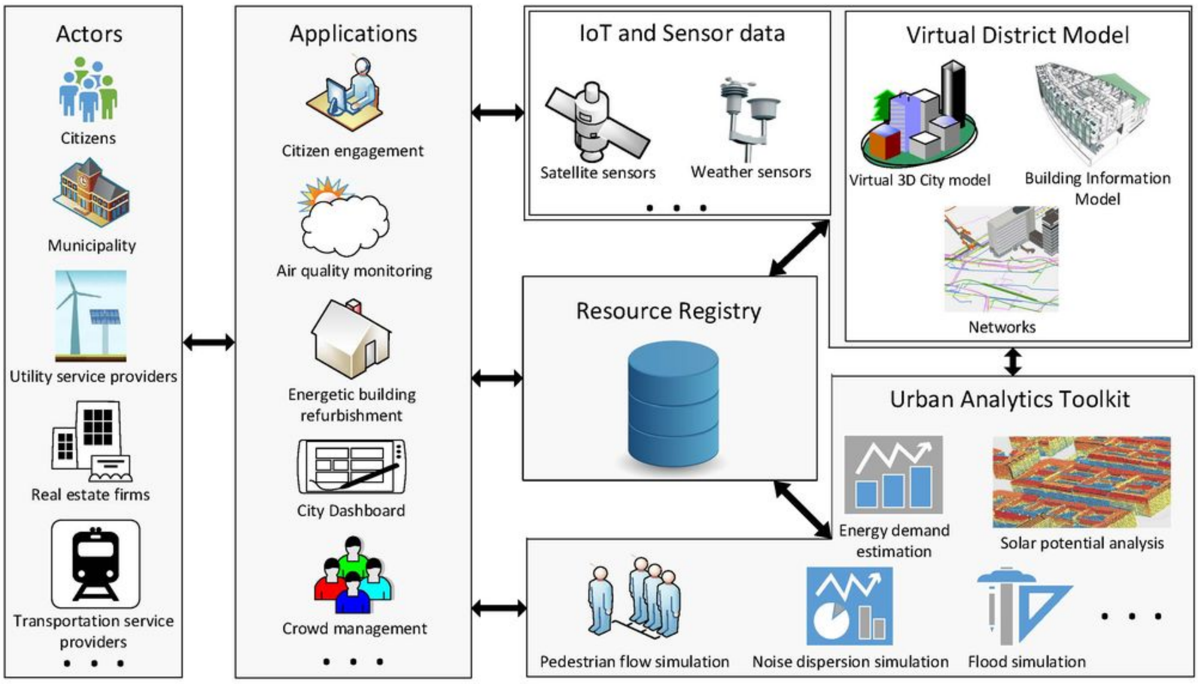
Conclusions on Data Integration

One cannot have a sustainable and smart district without proper handling of data integration. The term “Data Integration” does not only cover the challenges related to data, but also the modelling, structuring, managing and storage of the data, retrieving, translating, processing of the information, and handling and visualizing of the results.

In spite of the many existing technological solutions in the Smart City market, none of them is capable of dealing with all of the named problems so far. There is no doubt that Data Integration plays a leading role in resolving the problems. This is also being confirmed by many other organizations and communities who are working in the Smart City context such as TM Forum with their ‘City as a Platform’ Manifesto.

In the flagship project “Smart Sustainable District (SSD)” funded by EIT Climate-KIC, the “Data Integration” topic has been tackled within an individual working group. As the outcome of more than four years of the SSD Data Integration working group activities, the concept called “Smart District Data Infrastructure (SDDI)” has been developed. SDDI explicitly addresses Data Integration challenges in the light of distributed, heterogeneous multiple platforms. SDDI provides solutions to achieve both syntactic and semantic interoperability over the many distributed data resources from multiple stakeholders, tools, and applications on the basis of a range of mature international standards and already existing datasets.

The SDDI defines a structure which distinguishes between the different kinds of actors / stakeholders, the set of all applications, Sensors and IoT, virtual models of the physical environment, an urban analytical toolkit, and a central registry and catalog. All services, applications, analytical tools, sensors, and the catalog are implemented as distributed web services which are operated by different actors. The employed data models, exchange formats as well as the web service specifications are using Open Standards issued by the Open Geospatial Consortium, ISO, W3C, IEC, IETF, and OASIS (as described already above). The following picture illustrates the SDDI:



Starting from 2020 the SDDI concept is being employed within a number cities and rural areas in the state of Bavaria, Germany, in the context of the activities of the Bavarian Center for Digitization (ZD.B) within the platform "Smart Cities and Regions". A guideline on how to introduce SDDI in a city or region will be published by the end of 2020. Furthermore, many software components including the catalog system will be made available as Open Source software in the form of Docker containers that can immediately be deployed and operated in Cloud Computing environments. Further information: https://zentrum-digitalisierung.bayern/smart-regions/ (in German only)

More information about the SDDI can be found on these web pages (in English):

https://www.lrg.tum.de/en/gis/projects/smart-district-data-infrastructure/

http://www.opengeospatial.org/projects/initiatives/fcp1

Further readings:

Moshrefzadeh, M.; Chaturvedi, K.; Hijazi, I.; Donaubauer, A.; Kolbe, T. H. (2017): Integrating and Managing the Information for Smart Sustainable Districts - The Smart District Data Infrastructure (SDDI). In: Geoinformationssysteme 2017, Wichmann, downloadable from http://mediatum.ub.tum.de/node?id=1350813

Chaturvedi, K.; Kolbe, T. H. (2017): Future City Pilot 1 Engineering Report – Public Engineering Report. Open Geospatial Consortium, Doc. No. 16-098, downloadable from http://docs.opengeospatial.org/per/16-098.html

Chaturvedi, K.; Kolbe, T.H. (2019): Towards Establishing Cross-Platform Interoperability for Sensors in Smart Cities. Sensors 19(3). Freely downloadable from https://doi.org/10.3390/s19030562

Chaturvedi, K.; Matheus, A., Nguyen, S. H., Kolbe, T. H. (2019): Securing Spatial Data Infrastructures For Distributed Smart City Applications and Services. Future Generation Computer Systems 101. Freely downloadable from https://doi.org/10.1016/j.future.2019.07.002

Willenborg, B.; Sindram, M.; Kolbe, T. H. (2017): Applications of 3D City Models for a better understanding of the Built Environment. In: Behnisch, M.; Meinel, G. (Ed.): Trends in Spatial Analysis and Modelling. Springer International Publishing. https://doi.org/10.1007/978-3-319-52522-8_9

Kolbe, T. H. (2009): Representing and Exchanging 3D City Models with CityGML. In: Lee, J.; Zlatanova, S. (Ed.): 3D Geo-Information Sciences. Springer International. Downloadable from http://mediatum.ub.tum.de/doc/1145752/947446.pdf

Kutzner, T.; Chaturvedi, K.; Kolbe, T. H. (2020): CityGML 3.0: New Functions Open Up New Applications. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 19. Springer Publishing. Freely downloadable from http://dx.doi.org/10.1007/s41064-020-00095-z

Yao, Z.; Nagel, C.; Kunde, F.; Hudra, G.; Willkomm, P.; Donaubauer, A.; Adolphi, T.; Kolbe, T. H. (2018): 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML. Open Geospatial Data, Software and Standards 3 (5), Springer Publishing. Freely Downloadable from http://dx.doi.org/10.1186/s40965-018-0046-7

Applicaton of the SDDI methodology for digital agriculture:

Moshrefzadeh, M.; Machl, T.; Gackstetter, D.; Donaubauer, A.; Kolbe, T. H. (2020): Towards a Distributed Digital Twin of the Agricultural Landscape. Journal of Digital Landscape Architecture (5). Download from http://dx.doi.org/10.14627/537690019

Machl, T.; Donaubauer, A.; Kolbe, T. H. (2019): Planning Agricultural Core Road Networks based on a Digital Twin of the Cultivated Landscape. Journal of Digital Landscape Architecture (4). Download from http://dx.doi.org/10.14627/537663034