Fraunhofer
IAO

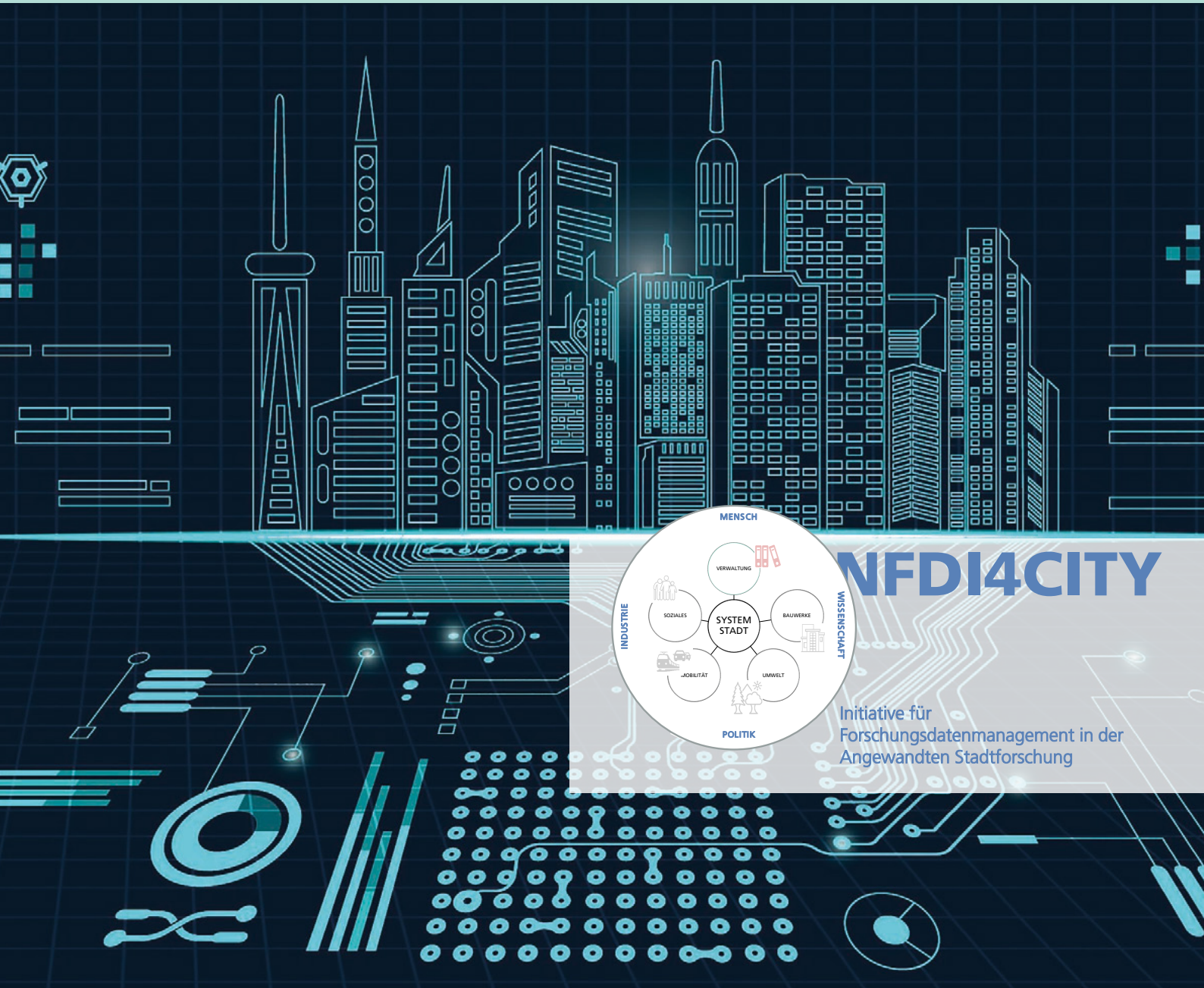
FRAUNHOFER-INSTITUT FÜR ARBEITSWIRTSCHAFT UND ORGANISATION IAO

2. FORUM URBANE DATEN, 7. BIS 8. OKTOBER 2020

WILHELM BAUER | OLIVER RIEDEL | STEFFEN BRAUN | EKATERINA DOBROKHOTOVA
JOHANNES SAUTTER (HRSG.)

FORUM URBANE DATEN 2020 – PAPERS

WORKSHOP OF THE NFDI4CITY INITIATIVE



Wilhelm Bauer | Oliver Riedel | Steffen Braun | Ekaterina Dobrokhotova
Johannes Sautter (Hrsg.)

2. FORUM URBANE DATEN, 7. BIS 8. OKTOBER 2020

FORUM URBANE DATEN 2020 – PAPERS

WORKSHOP OF THE NFDI4CITY INITIATIVE

Contents

Editorial: NFDI4City – Research Data Management for Applied Urban Research

Johannes Sautter [1], Steffen Braun[2] , Udo Lambrecht[3], Christoph Quix[4], Andrea Wuchner[5], Ulrike Küsters[5], Dorothea Iglezakis[6], Sonja Schimmler[7], Serafeim Alvanides[8]

Abstract: Within the agreement on the establishment and promotion of a National Research Data Infrastructure (NFDI), the Federal German Government and the Länder are striving for a discipline-specific service structure for research data, which aims for their long-term storage, safeguarding and accessibility. In the field of the applied urban research, data from different sectors such as administration, buildings, environment, mobility, social affairs must be considered and linked. In this workshop the NFDI4City-initiative met and discussed its overall vision, conceptual solutions as well as data domains inside and across these sectors. The goal of NFDI4City is to create new data-based overarching understandings of urban development through standardization and data analysis in the „Urban System".

Keywords: Urban research, Research Data, Data Management, Data Governance.

1 Introduction

Research Data Management including (meta) data standards, common data spaces, data governance, infrastructure and services is a topic of increasing relevance in the international research landscape. In addition to the scientific community, the public sector is therefore also in the spotlight when it comes to openness, availability and reuse of data. In order to be able to place urban phenomena, trends and developments in their overall context and to develop effective measures and solutions, the city – as well as applied urban research – must be understood as a holistic system whose sub-

[1] Fraunhofer IAO, Urban Systems Engineering, Nobelstr. 12, 70569 Stuttgart, Germany, Johannes.sautter@iao.fraunhofer.de, https://orcid.org/0000-0002-5934-7979

[2] Fraunhofer IAO, Head of Urban Systems Engineering, Nobelstr. 12, 70569 Stuttgart, Germany, Steffen.Braun@iao.fraunhofer.de

[3] ifeu Institut für Energie- und Umweltforschung Heidelberg gGmbH, Wilckensstr. 3, 69120 Heidelberg, Germany, udo.lambrecht@ifeu.de

[4] Fraunhofer FIT, Schloss Birlinghoven, 53757 Sankt Augustin, Germany, christoph.quix@fit.fraunhofer.de

[5] Fraunhofer IRB, Nobelstr. 12, 70569 Stuttgart, Germany, {vorname.nachname}@irb.fraunhofer.de

[6] University of Stuttgart, Universitätsbibliothek, FoKUS, Competence Center for Reseacrh Data, Holzgartenstraße 16, 70174 Stuttgart, Germany, dorothea.iglezakis@ub.uni-stuttgart.de

[7] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany, sonja.schimmler@fokus.fraunhofer.de

[8] GESIS Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln, Germany, S.Alvanides@gesis.org

areas influence each other. The discipline of applied urban research in context of the digital transformation of urban habitats towards smart cities unites different sectors, which are characterized by a homogeneous communication context (Scientific Community), a traditional body of knowledge, a common, well-founded research topic, a wide range of methods and scientific publication organs. The criteria of a scientific discipline are thus fulfilled. In contrast to many other NFDI consortia, however, existing basic research from other disciplines is used according to the principle of applied research. The following sections describe the data challenges, the vision as well as envisioned data management services of NFDI4City, before the data domains describe the researchers' point of view on their data challenges and structures.

2 Urban Development Questions and Data Challenges

In view of urbanization as a megatrend that will have a strong impact on the design of cities in the coming years, city data are of particular relevance to a wide range of stakeholders, such as politics, insurance companies and construction companies. The research data generated in applied urban research is made up of data from various sub-disciplines with their own standards. Looking at the individual sectors of administration, buildings, environment, mobility and social affairs (see Fig. 2), a picture of essential challenges emerges: (1) interdisciplinarity and therefore heterogeneity of research data, (2) lack of data availability due to separate data silos, (3) no comprehensive analyses of different data sets, (4) lack of methodological knowledge in data preparation and analysis, (5) outdated data, and (6) lack of data and process standards.

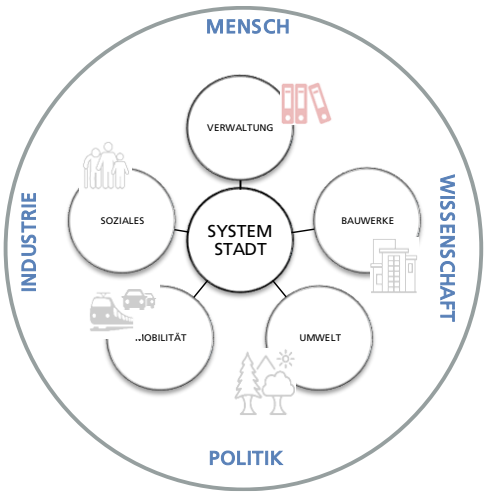


Figure 1: Applied Urban research as a discipline based on other disciplines

In urban research, heterogeneous and very individual data is created, where the difficulty lies in a limited possibility to standardize processes. At the same time, data protection concerns and economic exploitation interests in the data are also important factors in

shaping the necessary organizational regulations. In addition, cooperation is currently being established between the players – companies, politics, research and civil society – to ensure that data can generate added value.

Due to the character of an applied discipline, there are at least interfaces to the following other NFDI consortia[9]:

- *KonsortSWD (Sector Social Aspects)*
- NFDI4MobilTech (Sector Mobility)
- NFDI4Earth (several sectors)
- *NFDI4Ing (Sector Buildings)*

Analysis answering future interdisciplinary urban questions could be based on high-quality research data. Furthermore, the connectivity of city data to other disciplines for interdisciplinary research, such as climate research, could be guaranteed. For data producers, standardization will result in improved possibilities for exploitation, such as data portals that make different data available to end users. Quality standards and curation criteria also play a role in the long-term archiving and provision of data over many years. This task is performed by libraries and data centers. Defined quality standards and curation criteria can be used by them to assess and curate research data, thus supporting the long-term preservation and access to research data. Furthermore, specialists such as library staff and data stewards are enabled by generally accepted quality criteria to support scientists in achieving high data quality. In addition, the establishment of committees may create a cross-organizational framework in which the various actors in applied urban research and infrastructure operators can meet to develop quality standards. Table 1 shows some exemplary thematic areas that illustrate the interaction of the sectors in different urban research topics.

Topic	**Administration**	**Buildings**	**Environment**	**Mobility**	**Social Aspects**	**Resilience**
Urban land use planning (traffic and settlement areas, commercial areas, green spaces etc.)	x	x	x	x	x	
Social-spatial distribution of traffic-related air and noise pollution for the population	x	x	x	x	x	
Network planning and charging infrastructure in public and private buildings for electric mobility	x	x	x	x		x
Social inclusion and satisfaction of mobility needs of all population	x			x	x	

[9] www.nfdi.de , *oblique means funding started in 2020*

Topic	Administration	Buildings	Environment	Mobility	Social Aspects	Resilience
groups						
Social and climate-friendly living and building	x	x	x		x	x
City of short distances	x	x	x	x		
Public and private parking areas; parking space management		x	x	x		
Tourism: day guests/overnight stays, tourist flows	x			x	x	
Safety (traffic accidents, civil protection, crime...)	x	x		x	x	x

Table 1: Cross-sectoral topics of applied urban research

3 Vision

The initiative's aim is to create a new understanding for urban transformation and development via the intersectoral approach of the „Urban System". Adressing the ongoing acceleration of innovation cycles, technological progress, digital transformation and cultural change in public sector, valid, standardized and open research data can play a crucial role towards carbon-neutral, liveable and future-proof cities in the next decades.

In addition, access for partners in practice such as cities is just as much a goal as the interaction between science and society. The focus here is on creating homogeneous and interoperable data sets within the sectors and linking them through cross-sectoral standards. The required standards on both metadata and file level have to be developed and maintained within different sectors of applied urban research in coordination with the respective discipline-specific expert communities and NFDI consortia. With the help of organizational cross-sectoral data governance and technical support, research projects should realize a control and standardization of the comprehensive data management, which leads to an efficiency gain through research excellence as well as parallel and subsequent use. The interaction between science and society is also to be increased through citizen science, transparency of methodology and re-use vis-à-vis citizens and other data providers.

4 Envisioned Data Management Services

The initiative aims to achieve the mentioned goal of creating homogeneous data sets within the sectors of applied urban research by grouping files from researchers to data sets and link these to metadata. A major concept therefore are data domains, understood as logical grouping of research data collected with a certain method (e.g. questionnaires) or stored in a certain data format (e.g. sensor geodata with time reference). For metadata and file structures as well as for data handover and assessment processes cross-sectoral standards could be established and enhanced by sector and domain specific extensions.

Beside static data that is accessible in repositories, also dynamic data may play a role. Typical dynamic data are live-data from air-quality sensors that are available to researchers and practitioners by means of a public or restricted application programming interface (API). With the help of such technical solutions current research projects in urban research are to achieve efficiency gains through comprehensive data management, which in turn will lead to better data availability for practical use.

Organizational solutions including several roles for data stewardship as well as a workflow system complying to high standards in usability, e.g. based on the Data Governance Framework developed in the HEFE-project (Sautter, Wuchner 2020) are further envisioned services for researchers. The following services are envisioned along this vision:

- A data excellence organization, consisting of strategy, data governance, data management and a system architecture
- Data domains and responsible data owners within and across applied urban research sectors
- Common glossary and rules for metadata and files
- Transformation process for files along sensitivity levels from highly sensitive to open
- Phase oriented “Research process of the future”: Early categorization and curation of data records by service teams based on a set of rules, domain-specific document templates and metadata standards
- Overview for (planned) research data (depending on data value category) for scientific leaders and project managers

A central overall goal is to create a planning and results-based data counterpart of the scientific landscape that can be used for complex search queries as well as for analyses across different data sets including full text search and semantic integration.

5 Data Domains

In preparation of the workshop a “Call for domains” was tendered. With urban researchers as a target group, it asked for the data-challenge they face and their state-of-the-art in their

internal data management. The assessment or handing-over process was further an issue of interest as well as data types, meta data and file structures. Within and across sectors of urban research, highly relevant contributions on data domains were accepted as contributions to this workshop:

- **Construction Data**: Günter Wenzel from Fraunhofer IAO and Andrea Wuchner from Fraunhofer IRB describe the domain of building related data on the base of 3D geometry.
- **Urban Logistics Data**: Bert Leerkamp from the University of Wuppertal describes data related to transport demand and traffic flow with the aim to measure impacts on environment, urban structures and economy.
- **Trajectories:** As a important data type in the mobility sector, Daniel Krajzewicz, Alexander Sohr and Jan Weschke from DLR describe trajectories for modeling as well as monitoring traffic.
- **Urban Production Data:** Michael Hertwig, Adrian Barwasser, et al. from Fraunhofer IAO and others describe urban production data as information about input and output of production companies with the aim of urban symbiosis.
- **Crisis Management Field Exercise Data:** Johannes Sautter, Patrick Drews et al from Fraunhofer IAO describe quantitative and qualitative data assessed in field exercises wherein practitioners of civil protection move vehicles and equipment.
- **Geographic Information Systems and Data Challenges:** Christoph Sebald from Fraunhofer IAO describes solutions to geospatial data challenges.

Some of the described domains can directly be assigned to a particular urban research sector, such as trajectories to the mobility sector. Others, such as urban production or geodata at least relate to two or more sectors. Furthermore, not all sectors are represented in the described domains. In particular, social sciences are missing although they are very relevant.

6 Conclusion

A first small step towards the vision of creating a new data-based overarching understanding of urban development has been achieved by this description of the NFDI4City concept enhanced by data domains from a researcher‘s point of view. This will facilitate identifying commonalities across domains in the direction of a more precise and elaborated mission and roadmap of NFDI4City.

Contact

Steffen Braun, Steffen.Braun@iao.fraunhofer.de
Johannes Sautter, Johannes.sautter@iao.fraunhofer.de
http://s.fhg.de/nfdi4city

Bibliography

[Sautter, Wuchner, 2020] Sautter, Johannes; Wuchner, Andrea: Data Governance als Voraussetzung für Datenexzellenz an Forschungsorganisationen. Poster präsentiert auf der RDA Deutschland Tagung, Potsdam, 25.–27.02.2020. http://publica.fraunhofer.de/documents/N-586577.html

Domain Construction Data

Günter Wenzel [1], Andrea Wuchner [2]

Abstract: Located under the domain geodata, the sub-domain of construction (building) data deals with building-related data in the form of 3D geometry and component information. The data arises in the life cycle of a building from planning through execution and operation to deconstruction. A building can be understood as civil engineering and building construction.

Keywords: 3D Geometry, Building Information Modelling, Architecture, Building Data, Construction Data, Computer Aided Facility Management, Geo Data.

1 Introduction

A building (buildings, civil engineering) model is a collection of sub models which are generated or maintained along a building life cycle. In the planning or renovation phase they are generated in authoring systems by architects and specialist planners, in the construction phase by construction companies and trades and in the operation phase of the facility management. Construction Data belong to the domain of Geodata. Challenges are the following aspects:

- According to that we have highly divers data formats, standards and levels of detail. Nowadays we are far away from a common and current set of data for one building.
- Concerning construction data, the LOD (level of detail, level of development) depends on the kind of the model and is defined differently (e.g., countries vs. cities).
- There is no clearly defined ownership since the variety of different data-owners
- Keeping the data format current is nearly impossible due to many different authoring systems mostly with proprietary formats.

[1] Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Nobelstra□e 12, □0569 Stuttgart, Germany; guenter.wenzel@iao.fraunhofer.de

[2] Fraunhofer-Informationszentrum Raum und Bau IRB, Nobelstrae 12,0569 Stuttgart, Germany, andrea.wuchner@irb.fraunhofer.de, https://orcid.org/0000-0002-5549-3434

2 Method of data assessment or delivery

Construction data is bought or retrieved by external data sources, such as non-official land survey offices, expert or city planners, etc. Data of third parties are handed over with informal user agreements. Open source data and data retrieved from the research community is used as well.

In case there is no data to be retrieved by external sources, data will be modelled or collected during inventory control. A possibility to collect data is the inventory control by flying over private and public areas (with permission) using laser scanning or other scanning technologies. The bought, retrieved or collected data is stored in partial models and is screened and analysed according to the research project's guidelines.

This domain is characterized by a high variance of data types from many different sources and use cases along the building life cycle. Typical open standard formats are ifc (industrial foundation classes) for BIM which is often used in planning and construction but also in an increasing amount in the operation phase.

3 Data management

Until now, there is no standardised metadata in use for construction data.

The data is collected in partial models. The partial models may be integrated into a consolidated model following the terms of use. The software used is mainly in proprietary formats. The data is presented in systems such as Viewer, VR, AR, and XR systems. The data is currently stored on a research data server. In case of bigger datasets, the data is observed according to institute guidelines.

4 Conclusion

Planning and building is an essential part of the economy in many countries. In order to operate this at a high level and sustainable, it has to be continuously improved by the findings of research. The basis for this research is construction data. The management of construction data is not yet running optimally. There are still many challenges:

Interoperable standards for the continuation of the data within construction cycle must be defined and corresponding APIs set up. The data formats of the different trades must be compatible. Portals must be established in which building data are made openly available. Metadata standards need to be developed.

The construction community is called upon to face this challenge with everyone involved.

Urban Logistics Data

Bert Leerkamp[1]

Summary: The domain Urban Logistics Data covers data that describes urban logistics and aims at measuring impacts on environment, urban structures, economy and transport demand. Further relevant data are traffic flow data, transport behaviour data, data on vehicles and other mobile equipment used for pickup and delivery in cities as well as data on assets such as urban hubs and freight villages.

Keywords: urban logistics, transport data, traffic behaviour data, vehicle and asset data

1 Introduction

Urban areas fulfill basic functions as production sites, market places und transshipment points. The resulting commercial traffic flows are subject to urban transport and land use planning. An urban data base, which aims at describing the „urban organism" and provides information to govern it, should involve data about commercial goods and traffic flows.

„Urban logistics data" cover a wide range of information about logistic processes and logistic service providers, logistic assets and their functions, transport equipment, transshipment procedures (i.a. roadside delivery) traffic flows and environmental impacts.

„Sustainable Urban Logistic Plans" (SULP) are part of „Sustainable Urban Transport Plans" (SUMP) and contribute to integrated urban master plans. Quality standards for SULP include recommendations for data sets to describe the mode of operation and the impacts of urban logistic processes in order to develop them according to overriding targets of urban development (Agora 2020).

Furthermore, urban logistics data provide research in transport and urban planning. Methods to develop SULP and logistic action plans, research on economic and environmental impacts of measures and interactions between land use plans and commercial transport planning are essential elements of transport research. Modelling commercial transport in urban areas is still a challenging task as existing models for goods and traffic flows suffer from unsatisfying data bases and knowledge about

- transport demand of locations for production, trade and consumption,
- logistic processes in urban supply chains
- use of vans, trucks and bicycles in last mile logistics.

Therefore empircal studies and Big Data analysis should be considered as integral

[1] Bergische Universität Wuppertal, Pauluskirchstraße 7, 42285 Wuppertal, leerkamp@uni-wuppertal.de

elements of „Datasets4Cities“.

2 Data assessment and Data management

Main data categories for reserach and planning purposes are:

- transport demand data (by logistic market segments such as Parcel, food/non food, recycling... and/or recipients)
- traffic flow data (trucks and vans in urban areas),
- transport behaviour data (consolidation strategy; tour parameters such as stopps, tour length, time, ...),
- vehicles and other mobile equipment used for pickup and delivery in cities (vehicles: size, power units/e-mobility, ..., containers),
- assets (urban consolidation centres, urban hubs, freight villages, ...).

Transport demand data and behaviour are mostly provided by occasional queries carried out in research projects or other studies. Missing query and meta data standards complicate data use in research and planning and actually there is no common data base existing.

The German Federal Ministry for Transport provides data about commercial fleets (see KBA) and the use of commercial vehicles (see BMVI 2012). These data bases are open to public. Data about urban transsipment points can be investigated by internet based data mining. Retail data is partially available from local retail and consumer surveys. While aggregated population data is open to public there are still deficits in availability of detailed employment data, which are necessary to analyse origin and destination of goods flows.

3 Conclusion

Research in sustainable urban logistics actually suffers from heterogenious data structures and data providers. Developing standards for data structures, meta data requirements and organizing access for research and planning processes are essential tasks in the context of developing an urban research data infrastructure. A first approach could be an initiative for data mining and digitalization of administrative procedures in municipalities. Continious data production and data provision for urban research and planning affords ditigal administration processes with automatic data sampling according to research and planning demands.

Bibliography

[Agora 2020] Aichinger, Wolfgang und Leerkamp, Bert und Thiemermann, André und Wittenbrink, Paul (2020): Liefern ohne Lasten. Berlin: Agora Verkehrswende

[KBA] Kraftfahrtbundesamt (o.J.): Statistiken zu Fahrzeugbestand, Fahrzeugzulassungen und Güterver-kehr. www.kda.de

[BMVI] WVI, DLR, IVT, KBA (2012): Mobilitätsstudie „Kraftfahrzeugverkehr in Deutschland 2010“ (KiD 2010). Berlin: Bundesministerium für Verkehr und digitale Infrastruktur

Trajectories and Their Scientific Use

Daniel Krajzewicz[1], Alexander Sohr[2], Jan Volker Weschke[3]

Description: Mobility and traffic research relies on a large variety of data, used for both, modelling as well as monitoring traffic and its individual participants. One important data type are trajectories – positions over time, often enhanced by additional information – of single users or vehicles. While operating different systems to collect such data, data of this type are as well available from commercial suppliers. Different questions regarding the use of such data arise, which will be addressed in the following.

Keywords: scientific data usage, spatial and temporal superimposition.

1 Introduction

The research of mobility and subsequently generated traffic is a very data consuming task. One may find different methods for observing the behaviour of certain participants or the resulting traffic on our streets. They not only include different methods to be gathered, but differ as well in the spatial and temporal coverage. Conventional inductive sensors count the number of vehicles passing a road cross-section. Modern GPS application track persons or vehicles on their road and allow for determining the used modes of transport along a single trip. When coming to modelling traffic, the amount and diversity of needed data gets bigger. Data about the socio-demographics are needed to model the population of the regarded area, data about the available infrastructure and mobility offers is needed to model the usage of different modes of transport. Finally, information about usual behaviour, usually given as activity chains over a day, as well as information about available activity places is needed to compute the individual daily activities as well as for selecting the visited places.

One increasingly data type that is gaining in use are trajectories. A trajectory describes the movement of a person or a vehicle over time, in terms of a list of positions, often enhanced by additional information, such as the speed, the direction, the used mode of transport, etc. The DLR develops and operates own systems for collecting, storing, and processing trajectory data. To name are the projects KeepMoving and MovingLab. They deliver single user trajectories including the detection of the modes of transport used along a single trip.

[1] German Aerospace Center, Institute of Transportation Systems, Rutherfordstr. 2, 12489 Berlin, daniel.krajzewicz @dlr.de

[2] German Aerospace Center, Institute of Transportation Systems, Rudower Chaussee 7, 12489 Berlin, alexander.sohr@dlr.de

[3] German Aerospace Center, Institute of Transportation Systems, Rudower Chaussee 7, 12489 Berlin, jan.weschke@dlr.de

2 Data management and processing

Regarding the collection of trajectory data, we rely on an own communication chain that includes gathering data on smartphones or specific long-duration GPS trackers and sending them to own servers. For spatial data, we currently use GeoNetwork meta data servers together with Postgres databases with PostGIS extensions. Well-defined data may as well be found within the DLR traffic clearinghouse [Schulz, Brockfeld, Kelpin, Parnitzke, 2004].

The heterogeneity of the data used within mobility and traffic research raises the need for sophisticated methods for data retrieval, including possibilities to find data by their spatial coverage and the covered time. This counts as well for processing the data, where different methods for spatial and temporal combinations get necessary. Already available methods include geographic information systems (GIS) and spatial databases. Further processing methods cover topics such as plausibility checks on the data, filtering [Roth, 2018], map matching [Ebendt, Touko Tcheumadjeu, 2020], routing [Suske, Sohr, Neidhardt, 2020] as well as modern fusion and imputation [Heldt, Bahamonde Birke, Donoso, Heinrichs, 2018] methods and the use of regression models applied on derived measures.

3 Recognized issues

Some of the named data are available for free, yet it is usually necessary to include commercial data as well as administrative data. Such non-open data often comes with a time restriction for being used – after the end of the project, the data has to be deleted. Yet, on the other hand, good scientific practice raises the need to keep the data for being able to reproduce the obtained results.

4 Conclusion

Traffic and mobility research use highly heterogeneous data types, differing in spatial and time coverage. Besides the needs to develop methods for building catalogues and retrieving them easily, further data processing methods have been developed and keep being used.

Yet, some questions arise about how to deal with the trade off between a legal data usage within the framework of a scientific project while guaranteeing the reproducibility of the obtained results.

Bibliography

[Schulz, Brockfeld, Kelpin, Parnitzke, 2004] Schulz, Angelika und Brockfeld, Elmar und Kelpin, Rene und Parnitzke, Antje (2004) *Dateninformationsangebot der Clearingstelle für Verkehrsdaten und Verkehrsmodelle des DLR.* In: Frühjahrstagung 2003 - Verband Deutscher Städtestatistiker - Tagungsbericht. Frühjahrstagung 2003 - Verband Deutscher Städtestatistiker, ISBN 3-922421-43-1.

[Ebendt, Touko Tcheumadjeu, 2020] Ebendt, Rüdiger und Touko Tcheumadjeu, Louis Calvin (2020) *Toward Fast and Accurate Map-to-Map Matching of City Street Maps.* In: 2020 IEEE Intelligent Transportation Systems Conference, ITSC 2020, Seiten 3234-3239. 2020 IEEE Intelligent Transportation Systems Conference, ITSC 2020, ISBN 978-1-7281-4148-0.

[Heldt, Bahamonde Birke, Donoso, Heinrichs, 2018] Heldt, Benjamin und Bahamonde Birke, Francisco Jose und Donoso, Pedro und Heinrichs, Dirk (2018) *Estimating bid-auction models of residential location using census data with imputed household income.* Journal of Transport and Land Use, 11 (1), Seiten 1101-1123. University of Minnesota. DOI: http://dx.doi.org/10.5198/jtlu.2018.1040 ISSN 1938-7849.

[Suske, Sohr, Neidhardt, 2020] Suske, David und Sohr, Alexander und Neidhardt, Eric (2020) *Energy optimized routing for E-Vehicles.* In: 2020 International Conference on Smart Transportation and Future Mobility (CSTFM 2020). 2020 International Conference on Smart Transportation and Future Mobility

[Roth, 2018] Roth, Michael Helmut (2018) *Kartengestützte GNSS/IMU-Datenfusion mit Kalman-Filter und Co.: Anwendungen im Bahnbereich.* Positionierung und Navigation für Intelligente Transportsysteme (POSNAV 2018)

Urban Production – Value adding in urban environments

Michael Hertwig [1], Adrian Barwasser[2], Ivan Bogdanov[3], Xiaodu Hu[4], Joachim Lentes[5], Lara Waltersmann[6], Elmar Wendland[7]

Description (same as in form): Productions are an important aspect of a city framework. As stakeholder with a big impact on sustainability a radical reduction of emissions and resource consumption is required. By setting up symbiosis and interdependencies it is possible to reduce negative impacts and increase improvements in collaboration. To evaluate potential symbiosis measures, it is required to have information about input and output of production companies together with further information. This data needs to be collected, stored and applied to the framework of ultra-efficiency of industrial estates.

Keywords: Ultra-efficiency data; urban production data; urban symbiosis; company-specific data

1 Introduction

Production is an important aspect for urban areas. Factories offer employment for many citizens. Additionally, the society can profit from close collaboration with enterprises. Since many people are concentrated in a limited area several services could be offered with a high quality based on higher demand.

On the other hand, factories are one of the biggest emitters of greenhouse gases, pollution and other negative impacts. Current efforts are focussing on increasing sustainability in factories by digitisation [L16]. Analysis showed that this is the right track. However, limiting the area of focus to a factory is also limiting the potential for improvement [SH16].

The Project "Ultraeffizienz4Industriegebiete" is focussing on industrial park and

[1] Fraunhofer Institute for Industrial Engineering IAO, Digital Engineering, Nobelstrasse 12, Stuttgart, 70569, Michael.hertwig@iao.fraunhofer.de, https://orcid.org/0000-0003-0534-098X

[2] University of Stuttgart. Institute for Human-factors and Technology Management IAT, Nobelstrasse 12, Stuttgart, 70569, Adrian.barwasser@iat.uni-stuttgart.de

[3] Fraunhofer Institute for Industrial Engineering IAO, Energy Systems, Nobelstrasse 12, Stuttgart, 70569, ivan.bogdanov@ipa.fraunhofer.de

[4] University of Stuttgart. Institute for Human-factors and Technology Management IAT, Nobelstrasse 12, Stuttgart, 70569, Xiaodu.hu@iat.uni-stuttgart.de

[5] Fraunhofer Institute for Industrial Engineering IAO, Digital Engineering, Nobelstrasse 12, Stuttgart, 70569, joachim.lentes@iao.fraunhofer.de

[6] Fraunhofer Institute for Industrial Engineering IAO, Digital Engineering, Nobelstrasse 12, Stuttgart, 70569, lara.waltersmann@ipa.fraunhofer.de

[7] Wirtschaftsförderung und Standortentwicklung Rheinfelden (Baden) GmbH, Karl-Fürstenberg-Straße 17, Rheinfelden (Baden), 79618, e.wendland@wst-rheinfelden.de

industrial estates. By implementing measures of ultra-efficiency on an urban level the project team hopes to increase the sustainability potential tremendously.

2 Method of data assessment or delivery

The objective is to set up a framework of symbiosis between stakeholders in the urban area. Symbiosis with relevance are exchange of energy or material which means by-product of one stakeholder could become an input of other stakeholders. Additionally, it is also possible to share infrastructure or service to reduce doubled effort and increase the performance quality due to better capacity utilisation [BBSR16].

Based on the concept of ultra-efficiency there are 5 categories for data – material, energy, emissions, human/staff and organisation [LM16]. In an ongoing project measure of symbiosis are detected in all 5 categories (see in Figure 1, solid lines). Additional measures to increase the collaboration between the stakeholders have be identified based on collected data (see in Figure 1, dashed line).

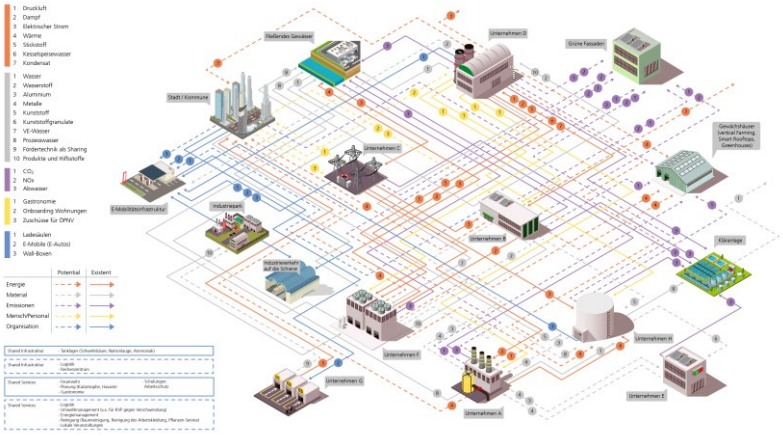


Figure 1: Framework of existing and potential symbiosis effects

The data was collected by use of a questionnaire. All identified stakeholder got a simple collection table which helped them to put together all required data. To support the stakeholder an expert was explaining the questionnaire and its objective in the beginning.

Based on the collected data a table for each stakeholder could be provided (see Figure 2). The data is sorted to the 5 categories [LH19]. Each data set is consisting of type and amount. Additionally, the data set could contain source and sink to identify it as input or output. By adding additional information some aspects of the data set can be analysed in detail.

For a long-term improvement also the ultra-efficiency measures need a description. It would be important to provide information about the measure itself and potentials for improvement [P18]. The decision to implement measures will be performed by the representatives of the local stakeholders. This requires a well-documented impact analysis. Based on the analysis the stakeholder can decide about potential investment or evolving efforts. This information is currently not standardized but already summarised in ultra-efficiency fact sheets. This information is collected by domain experts analysing current research papers and sales offers form distributors.

	Type	Amount	Source	Sink	Additional information
Material	Steel	70 tons	Supplier	Company	
	Wood				
	Hydrochloric acid	750.000 liters	Company	network partner	34% of the partner company's hydrochloric acid requirements are covered
	Hydrogen	1.870.000 liters	network partner	Company	43% of the company's hydrogen needs are covered
	Packaging	8.500 tons	surrounding area, roughly 2. Mio people	Company	
	Steel chips	15 tons	Company	residual recyclers	
	Steel waste	3 tons	Company	residual recyclers	90 % residual recycler, 10 % workshop for training
Energy	Electricity	1,2 MWh	Electricity provider	Company	
	Thermal energy	70 MWh	District heating supplier	Company	
	Thermal energy	-	Company	Umwelt	Waste heat at 35 °C is not reused
	Thermal energy	20 MWh	Company	City	80% of waste heat at 75 °C is delivered to the city
	Compressed Air	70.000 Nm³	own compressors	Company	
	Compressed Air	10.000 Nm³	Compressors, network partner	Company	12.5% of the company's compressed air requirements are covered
Emission	CO_2	-	Company	Environment	Curing fumes through chimney, filter in use
	NOx-Emission	10 t	Company	Environment	Exhaust gases through chimney, filter technology in use
	Waste Water	350.000 m³	Company	Waste water treatment plant of the city	
	Process water	150.000 m³	Company	Chemical waste water treatment plant	
	Noise	70 dB	Company	Environment	Sound insulation of machines - 25% noise reduction
Human / Staff	employee - car	120 von 500	Home	Company	
	employee - public transport	350 von 500	City	Company	bus connection to city center
	employee - bicycle use	30 von 500			5 of it by e-bike
	Illness days	12			2% work accident, 12% commuting accident, 80% illness (flu etc.), 6% physical complaints
	Training and qualification of employees	50 h	Lean Experts / IT Experts / etc.	Employee	
	Gastronomie am Standort	100.000 visitors			
Organisation	Consultancy service	5 Days	Consultancy office	Company	
	external logistic servicee	250.000 €	Logistic provider	Company	
	Anzahl Feuerwehreinsätze	1	Fire brigade city		
	Network / esxchange	4 events	-	-	2 events at the site, 2 events at the city's business development department
	online material supervision	15-Min.-Intervalle			
	e-mobility infrastructure	10 cars, 5 charging solutions			Extension by 50% within next 2 Years

Figure 2: Example of collected data from local stakeholders

3 Data management

The data of the stakeholders is currently use for a manual analysis. Based on that, the data is collected in a table scheme as provide by MS Excel. In the further research work the data shall be used in demonstrator unit. This demonstrator unit is in development currently. In future the data shall be stored in a format supporting different data processing. From the current perspective a machine-readable format is suitable. Because

in future the data still results from questionnaires or interviews with local stakeholders. Depending on the total number of data sets, it will be high effort to put all data set to work in the demonstrator unit.

The meta data connected to each data set contain information about the stakeholder. The meta data set includes information about size, location, representing stakeholder, and other stakeholder specific information. Due to the fact, that some of this information allows conclusions to the stakeholder the data muss be classified. In addition, the meta data contains a connection to the procedure of data acquisition.

In the future, there will different data sets connected to a stakeholder. This results from different analysis frames and alternatives which shall be discussed by the local stakeholders. All this information shall be stored in a standardized format. As machine readable format is allows the utilization in the demonstrator unit.

4 Conclusion

Currently, the project collected a first data set of 30 stakeholder. Due to fact that the data acquisition was performed for a manual analysis the existing data sets are not complete regarding all indicators. However, it is a first step an will be completed within the next year for a single location. The developed qway of proceeding is supporting the data collection in future activities improving collaboration of industrial companies leading towards stronger sustainability.

5 Acknowlegdement

We gratefully acknowledge the support of the Baden-Wuerttemberg Ministry for Environment, Climate and Energy within the Project “Die Ultraeffizienzfabrik – Symbiotisch-verlustfreie Produktion im urbanen Umfeld” and “Ultraeffizienzfabrik - .

Bibliography

[LM17] Lentes, J., Mandel, J., Schliessmann, U., Blach, R., Hertwig, M., & Kuhlmann, T. (2017). Competitive and sustainable manufacturing by means of ultra-efficient factories in urban surroundings. International Journal of Production Research, 55(2), 480-491.

[LH19] Lentes, J., & Hertwig, M. (2019). Towards Ultra-Efficient Industrial Areas. Procedia Manufacturing, 39, 804-813.

[L16] Lentes, J. (2016): Urbane Produktion. In: D. Spath, E. Westkämper (ed.), Handbuch Unternehmensorganisation, Springer Reference Technik, Springer Berlin Heidelberg.

[SH16] Spath, D.; Hertwig, M. (2016): Konzeptstudie Holistische Stand-ortentwicklung von produzierenden Unternehmen unter Berücksichtigung von Wechselwirkungen mit dem Umfeld, University of Stuttgart.

[BBSR16] Bundesinstitut für Bau-, Stadt- und Raumforschung (2016): Nachhaltige Weiterentwicklung von Gewerbegebieten, ExWoSt-Information 49/1, Bundesamt für Bauwesen und Raumordnung, Berlin.

[P18] Pichlmeier, F. (2018): Ressourceneffizienzpotenziale von Gewerbegebieten. VDI ZRE Publikationen: Kurzanalyse Nr. 22. VDI Zentrum Ressourceneffizienz GmbH (VDI ZRE), Berlin.

Crisis Management Field Exercise Data

Johannes Sautter [1,2], Patrick Drews [1]

Summary: This abstract describes the domain of crisis management field exercise data. As one of several possible data formats, patient actor data from mass casualty exercises are described. They are asses by patient actors in the field and stored in a table spreadsheet. As one analysis method, quality indicators can serve as a an aggregated view with the special purpose of evaluating the exercise actor's performance in crisis management and patient relief.

Keywords: Exercise data, mass casualty exercises, quantitative data, quality indicators.

1 Introduction

Crisis simulations and exercises are the only chance first responders and technical relief practitioners have to gain feedback on their mission skills. Contrary to serious games and simulation exercises, in field exercises, practitioners move real vehicles and equipment to accomplish and enable individual and organizational learning for practitioners. In general, there is a lack of research in the area (Beerens et al 2016). Regarding exercise data, few data sets are available publicly.

Depending on used evaluation methods, field exercises generate various kinds of data. As **data-challenge** this domain describes both qualitative data (e.g. audio/video media files, observer reports) and quantitative data (e.g. timestamps with achieved accomplishment milestones, patient treatment data). The data domain of field exercise data belongs to the sector resilience that itself is of cross-sectoral nature related to classical city sectors like mobility, environment and administration.

2 Method of data assessment or delivery

Field exercise data is assessed using different **exercise assessment methods**, such as process evaluation and qualitative observation (Drews et al 2019). Additionally, to assess data, handed-over mission-operational artefacts such as mission reports could in the aftermath serve as exercise data and resource for evaluation.

[1] Fraunhofer IAO, Urban Systems Engineering, Nobelstr. 12, 70569 Stuttgart, Germany, prename.surname@iao.fraunhofer.de

[2] https://orcid.org/0000-0002-5934-7979

One established **quantitative assessment method** is regarded as mature with regard to process experience, data quality and awaited measurement errors (Sautter et al 2014, Drews et al 2016, Sautter et al 2019). Qualitative approaches are also quite mature as well as established among practitioners and researchers. Other quantitative methods require some further development (Drews et al 2019).

The domain consists of described **various data types**. As an example, this abstract refers to quantitative patient actor data, assessed by the first aforementioned method.

3 Data management

The dataset structure is adopted from a generic framework for Research Data Governance and consists of bibliographic metadata and files as major categorization (Sautter, Wuchner, 2020). Beside datacite and dublincore[4] fields the schema contains also more specific fields like "research question" and an own entity for "research method".

As a possible metadata structure, an **exercise data set** could be described by a title, such as "Two mass casualty field exercises run with bus scenario in order to evaluate electronic patient tags' efficiency". A further metadata field could describe the exercise setting as well as used research methods and obtained data in 5 sentences. The metadata field research question could in this example be "Are electronic patient tags more efficient then manual patient tags?". The research object could be described as "Mission accomplishment of first response and fire brigade units". The research subject could be outlined as "Patient actors assessing particular mission accomplishment milestones".

As the method has further been implemented by a special exercise evaluation spreadsheet, **a "research instrument dataset"** could be described as follows. The name of the research instrument that could be "mass casualty exercise evaluation tool" in this case as well as an adequate five sentences description. Important thereby is the version of the research instrument due to potential schema evolution. A link to the research instrument website could be given, which may in this case be the reference to a corresponding publication offering the table spreadsheet as a working tool "manv-uebungen.iao.fraunhofer.de".

For describing the **methods** that have been both applied for data assessment and data evaluation, methods could be described by their title, type description and literature reference. Table 1 illustrates such a range of methods for the example dataset.

[4] Bibliographic meta data standards

Research-method-ID*	Research method	Type of research method	Description of research method	Bibliography
1	treatment assessment sheet[5]	data collection	form in which role players and/or evaluators document time stamps and essential interventions and result of triage.	(Sautter et al 2014), (DRK 2016), (Drews et al 2018), (Sautter et al 2019)
2	observer protocol	data collection	Semi-structured report, which is filled out by observers.	(DRK 2016), (Pottebaum et al 2014)
3	Camera data	data collection	Raw data of cameras used to capture the exercise.	(Pottebaum et al 2014), (DRK 2016)
4	Speech protocols	data collection	Protocols of given orders and reports within an exercise	(Pottebaum et al 2014)
5	quality indicator calculation[6]	data analysis	Calculation of quantitative indicators based on process-milestones collected by patient actors	(Sautter et al 2014), (DRK 2016), (Drews et al 2018), (Sautter et al 2019)
6	summary/ synopsis of observer protocols	data analysis	Experts observing the exercise create a summary of their observations	(DRK 2016)
7	collected presentation cards	data collection	Trainees as well as patient actors collect 3 positive and 3 negative aspects on cards	(DRK 2016)

Table 1: Example methods, potentially relevant for a field exercise dataset

[5] German Versorgungsbewertungskarte, see at http://manv-uebungen.iao.fraunhofer.de/
[6] Based on a spreadsheet, e.g. available at http://manv-uebungen.iao.fraunhofer.de/

Additionally, domain-specific metadata fields could allow a first overview on field exercise dataset for domain experts. Possible **domain-specific meta data** could look like illustrated in table 2 (first approach).

Property	**Value**	**Unit**
Situation	Train crashed to car, 24 injured	Text
Mass casualty	yes	Boolean (yes/no)
Weather	Cloudy, light rain	Keywords
Mission units	LF, RTW, RTH, KTW, KOM[7]	Keywords
Day time	1.20 AM, night, weekend	Keywords

Table 2 : First approach on domain specific meta data for field exercises

A typical **grouping criteria for field exercise data** is the scope of a particular experiment/exercise or exercise series with a particular focus in terms of exercise goal and research question. Therein typically one to three exercise runs could be part of a single dataset. In case more than one exercise run is grouped by the dataset, files would be grouped according to exercise runs on the top level and according to assessment or evaluation methods on the second level of the file system structure.

Currently, exercise **data is stored** on local file systems. Metadata are not yet assessed and maintained systematically. As **raw data** files generated by assessment methods, analog data, such as hand-written papers, may play a role. After their digitization and assessment, for the quantitative patient actor data, the aforementioned spreadsheet serves as a container file. Further media data such as helmet camera movies potentially enhances the data per exercise run. Observer reports also may be digitized and brought to a common format. An adequate option for observer reports e.g. could be to not digitize the handwritten paper reports but just ask each observer to provide a 1-pager containing his main observations as **aggregated data** after the exercise run.

4 Conclusion

Field crisis management exercises are both important for practitioners and essential as a research field. An assessment and usage of exercise data for both mission debriefing and scientific insights has been achieved so far. This domain abstract envisions an open standard for exercises data, containing out of metadata and file structure containing data from qualitative and quantitative exercise assessment methods. The granularity of the

[7] German abreviations for fire brigade vehicles, ambualance, helicopter, bus

data set is dominated by the corresponding research question, while a single data set contains various field exercise runs, that represent practitioners running a mission tactic in a simulated crisis environment e.g. represented by patient actors.

As a next step, dataset from particular exercises shall be published on a research data repository. Due to privacy and data protection limitations, just an excerpt of the whole data structure can most likely be made openly accessible. According to the vision of NFDI4City and the preceding HEFE project, an internal pendant of the dataset could nevertheless contain comprehensive data that has not been deleted in the data protection compliant data extraction and deletion process.

As remaining task to the research community, a classification e.g. for all possible crisis scenarios is needed to be established as a standard for exercise dataset.

Bibliography

[Beerens et al, 2016] Beerens, Ralf; Josef Johanna; Tehler, Henrik: Scoping the field of disaster exercise evaluation - A literature overview and analysis. In: *International Journal of Disaster Risk Reduction* vol. 19 (2016), pp. 413–446, DOI: 10.1016/j.ijdrr.2016.09.001

[Drews et al, 2018] Drews, Patrick; Sautter, Johannes; Offterdinger, M.; Rieger, A. (2016): Using quality indicators to evaluate MCI exercises. Experiences from an airport exercise, Poster presented at IPRED 2018, 5th International Conference on Healthcare System Preparedness and Response to Emergencies and Disasters, 14-17 January 2018, Tel Aviv, Israel. http://publica.fraunhofer.de/documents/N-497729.html

[Drews et al, 2019] Drews, Patrick; Sautter, Johannes; Kübel, Marcel (2019): Realitätsnah oder genau? Triangulative Übungsevaluation, Erfahrungen aus dem REBEKA-Projekt, Kröling, Sophie (Hrsg.): Konzepte zur Steigerung der Resilienz von Einsatzkräften. Ergebnisse aus dem Forschungsverbund REBEKA, (Schriftenreihe Sicherheit 26), ISBN: 978-3-96110-237-2 (Print), ISBN: 978-3-96110-236-5 http://publica.fraunhofer.de/documents/N-561803.html (Online), pp.175-191, Berlin: Freie Universität Berlin, 2019.

[DRK, 2016] Deutsches Rotes Kreuz (2016): Durchführung und Auswertung von MANV-Übungen, Wissenschaftliche Erkenntnisse und Best Practices, Beteiligt: Sautter, Johannes; Schneider, Friederike; Kippnich, Uwe; Wirth, Sven; Habermann, Manuel; Siebel, Claudia; Kippnich, Max; Frings, Sandra; Häusler, Hardy; Max, Matthias, Berlin: Deutsches Rotes Kreuz, 84 pp., Schriften der Sicherheitsforschung, 3, 2016. http://manv-uebungen.iao.fraunhofer.de/

[Pottebaum et al, 2014] Pottebaum, Jens; Marterer, Robin; Schneider, Steffen (2014): Taxonomy of IT support for training emergency response & management. In: Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014 S.R. Hiltz, M.S. Pfaff, L. Plotnick, and P.C. Shih, eds. http://idl.iscram.org/files/pottebaum/2014/848_Pottebaum_etal2014.pdf

[Sautter et al, 2014] Sautter, Johannes; Habermann, Manuel; Frings, Sandra; Schneider, Frederike; Schneider, Bernhard; Bracker, Holger (2014): Übungsunterstützung für Einsatztrainings des Massenanfalls von Verletzten (MANV), in: Plödereder, E.; Grunske, L.; Schneider, E.; Ull, D. (Hrsg.); Gesellschaft für Informatik -GI-, Bonn: Informatik 2014. Big Data - Komplexität meistern. CD-ROM: Tagung der Gesellschaft für Informatik, 22. - 26. September 2014 in Stuttgart, Deutschland, Bonn: Köllen, 2014 (GI-Edition - Lecture Notes in Informatics (LNI). Proceedings 232), ISBN: 978-3-88579-626-8. http://publica.fraunhofer.de/documents/N-323651.html

[Sautter et al 2019] Sautter, Johannes; Offterdinger, Matthias; Drews, Patrick (2019): Qualitätsindikatoren für den Massenanfall von Verletzten (MANV) für Ausbildung, Übung und Einsatz, Poster präsentiert auf dem 19. Kongress der Deutschen Interdisziplinären Vereinigung für Intensiv- und Notfallmedizin; Kooperation führt zu Entwicklung; 04 .- 06. 12.2019, Hamburg, http://publica.fraunhofer.de/documents/N-578051.html

[Sautter, Wuchner 2020] Sautter, Johannes; Wuchner, Andrea (2020): Data Governance als Voraussetzung für Datenexzellenz an Forschungsorganisationen. Poster präsentiert auf der RDA Deutschland Tagung, Potsdam, 25.–27.02.2020. http://publica.fraunhofer.de/documents/N-586577.html

Geographic Information Systems and Data Challenges

Christoph Sebald [1]

Abstract: In order to empower cities and their decision makers, to enhance the productivity across their organisations and institutions, it remains of utmost importance to organise and manage all available resources in the most efficient way. GIS applications management tools help and play a vital role in often building the centre around geo-spatial data and an institutional information infrastructure, support to enhance workflows to better collect, store, maintain, prepare, and share the data available.

Keywords: Geospatial data, big data, data format, database, geographic information system, data steam and data flow, complex data, geo-visualization and visual analytics.

1 Introduction

Organizations, Institutions, and Departments, they all have data users or data consumers with a variety of distinct roles and needs. There is a constant increase of data at their disposal, consequently working or consuming more data than ever before, as the world appears to become even more complex with more data being present.

Main challenges are to treat data, or big data as an important asset and to becoming more data-oriented or data-driven is of central und utmost importance, disregarding of whether the goal is to achieve digital transformation or to develop a new and better way of deriving knowledge through situational insights and awareness for informed decision-making (Robinson at al 2017). Successful data management as an essential prerequisite due to these vast amounts of data of different types form different sources.

Geospatial data or big data, and its formats range from simple raster or image formats to rather complex database structures and formats are designed to describe, and frequently visualise features placed in specific geographic locations in space and time. Consequently, the data functionality is supported by various geospatial formats that can help represents the degree to which geographic information systems and other computer systems can create, use and modify the geospatial resources, and will be described within the geospatial format descriptions. The higher the quality of the data, the better its representation.

There are also real-time sensor data, environmental data, customer or population data, aerial and satellite imagery, and so on. All data and information that can or must be stored, rearranged, aggregated and processed to derive new formats, or derive specific analytical insights and knowledge. The main challenge is in finding the right set of tools that enable

[1] Fraunhofer IAO, Urban Systems Engineering, Nobelstr. 12, 70569 Stuttgart, Germany, christoph.sebald@iao.fraunhofer.de

users to efficiently generate value through the process of handling that data (garbage in – garbage out), aiming to deriving these specific, meaningful insights and outputs and converting new knowledge into action, i.e., better decision-making.

Distinct challenges are also the lack of publicly available or affordable data in some research projects. Why? First, because commercial data or data licenses are costly and the more details or accuracy is needed, especially in advanced processed datasets, the higher the cost for acquisition or use licensing. The same is true with historic, real-time, near real-time, future models and data and representations. Second, the use of Open Source despite a lack of detail and validity on specific datasets. Open Source is great, however, there is a risk for biases. Nonetheless, having little data to work with is usually better than having no data at all. Third, resource intense and time-consuming processing on heavy, large open source, sensor or satellite datasets. Finally, there are data domain specific and other related challenges, such as formatting issues due to multiple data types that need to be addressed and considered.

2 Method of data assessment or delivery

Data is usually retrieved from either commercial data suppliers or Open Source. The Open Source option may entail countless online sources, mostly for larger scale, greater than city district, city or county boundaries. In comparison, there are many companies that act as data suppliers, each providing the same or a similar data portfolio to choose from. With any data set there are specific formal or informal user agreements. Other data might be retrieved from within the research community if possible.

Another challenge is how web-based GIS development and data visualisation is delivered, and how much efficiency we can get out of delivering the data a certain way. The main obstacle is that it is rather difficult, often requires code to be written. This requires a human to solve, instead of a machine.

Then, data assessment must be conducted with strong computational power (e.g., processor capacity, large memory, graphic, reliable and enduring components of hardware to process heavy data (several GBs), and so on. This also entails regular servicing Inspection, projection, enfrichement, manipulation, extraction, conversion, assimilation

As mentioned before, there are multiple data types and the geospatial domain is characterized by a large variety of different data, formats or file types. There are, for instance, vector GIS file formats, raster GIS file formats, compressed raster file formats, geographic database file formats, Relational Database Management System (RDBMS) Enterprise, LiDAR file formats, CAD file formats, elevation file formats, web file formats, multitemporal file formats, and many more.

3 Data management

Metadata is important and as valuable as data itself, but how are metadata described? First, it is data that provides information about other data. Second, the metadata summarises basic information about data, making finding and working with specific instances of data easier. Third, the metadata can be created manually to be more accurate, or automatically and contain more basic information.

- Metadata is information about data, like a library catalogue record, metadata records document the who, what, when, where, how, and why of a data resource.
- Geospatial metadata describes maps, Geographic Information Systems (GIS) files, imagery, and other location-based data resources.
- Because spatial data is the fuel of a GIS, it is important to know if the data will meet user needs.
- Data users need metadata to locate appropriate data sets or their specific parameters within.
- Metadata provides information about the data available within an organization or from external services or other external sources.

All important information and specific characteristics are collected, recorded and stored in metadata that should or must include details about its features, geographic coordinates, projection, scale, datum, and more. Unfortunately, this is not always the case.

Any data or data source, disregarding of its origin should provide or include information that would allow the user or viewer to determine how accurate the specific features displayed corresponded to its geographic "reality," or determine its accurate location from the data source and its handler.

What are potential criteria for grouping various data in a data set/data frame/item etc.? Data could and should be grouped by relationship, location, source, or by context / topic, the source or origin. The level of quality of the data, potential product that can be derived from, collected and stored in (geospatial) models or specific databases.

4 Conclusion

There are many ways to access and use various datasets, through either commercial providers or open source. The quality of the metadata mostly depends on the source derived or provided. It often also determines the cost, or should we say the financial burden and legal options or restrictions about its use, dissemination and allowed levels of data fusion and further manipulation.

The world wide web is packed with guidelines about data specification, specific information sources, both free and payed to learn more about "big GIS Data", disregarding of the domain. Efforts must focus and take place to define potential linkages across research goals in the geospatial domain, in order to tackle complex data challenges.

Coordinated workshops and web-based collaborative efforts could help to educate people and experts to pursue challenges with the geospatial big data challenge of the 21st century, for better data to be available to the scientific community and the public alike.

Bibliography

[Robinson et al 2017] Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., ... Sluter, C. R. (2017). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. International Journal of Cartography, 3(sup1), 32-60. https://doi.org/10.1080/23729333.2016.1278151

Impressum

Kontaktadresse:
Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Nobelstraße 12, 70569 Stuttgart
www.iao.fraunhofer.de

Johannes Sautter
Telefon +49 711 970-2387
Johannes.Sautter@iao.fraunhofer.de

NFDI4CITY – Initiative für Forschungsdatenmanagement in der Angewandten Stadtforschung
http://s.fhg.de/nfdi4city

urn:nbn:de:0011-n-6088184
http://publica.fraunhofer.de/dokumente/N-608818.html

Titelbild: © Jackie Niam – Adobe Stock

Ebenfalls verfügbar:

Forum Urbane Daten 2020 – Vorträge

urn:nbn:de:0011-n-6088192

http://publica.fraunhofer.de/dokumente/N-608819.html