
Map Production Supporting Activities of Digitalization and AI

Markus Jobst^{a,b} *, Georg Gartner^a

^a *Research Group Cartography, Vienna University of Technology, markus@jobstmedia.at; Georg Gartner, georg.gartner@tuwien.ac.at*

^b *Austrian Federal Office of Metrology and Surveying, markus.jobst@bev.gv.at*

* Corresponding author

Abstract: The notion “digitalization” comes into vogue in many corners of information and knowledge technologies. Even politicians are aware of the power of digitalization and its impact on society. In general digitalization is not directly dealing with webmapping or location based services, although digital and web-based procedures are the key. But what else could these initiatives of digitalization mean for the mapping domain? Is there an impact on map production, geoinformation management and the way we provide and use geospatial knowledge? The aim of this contribution is to formally extend the function of cartography by the aspect of artificial knowledge expansion and therefore to highlight the basics, requirements and additional emerging methods in map production.

Keywords: Map Production, Service-Oriented Mapping, Dataintegration, Digitalization, Artificial Intelligence, Geovisualization

1. Introduction

Digitalization enables smart implementations and prepares artificial intelligence. In general it is not directly dealing with webmapping or location based services, although digital and web-based procedures are the key. But what else could these initiatives of digitalization mean for the mapping domain? Is there an impact on map production, geoinformation management and the way we provide and use geospatial knowledge? This contribution sets the scene for modern map production and explores possible influences of digitalization on service-oriented mapping and geoinformation management.

2. Confrontations in map production

Map production is confronted with an increasing number of data sources, a variety of use cases and continuously changing environments for map production. Therefore, a service-oriented environment seems to be an appropriate basis.

The opening of earth observation data, spatial in-situ recordings and scientific data increase the number of data sources (Juckes 2016). In the meantime, these sources are

generally not restricted by thematic domains or boundaries, but may focus on specific aspects of the research topics. Some topics are narrowly focused in space, whereas others extend coverage and thematic linking as part of their research (Wilkinson 2016). Due to an increasing demand of environmental and global observation, in those datasets the coverages are increased as well. In that case any map production is confronted with the selection of appropriate sources, where “appropriate” concerns reliance, actuality, precision, completeness and consistency for planned use cases.

Following the basic idea of graphically transmitting spatial knowledge via maps, its use cases depend on transmission media characteristics (serving the human perception system) and situation- and task induced requirements. Even if maps are produced for specific situations or tasks, their perceptual encoding (graphical variables) needs to be adopted for the transmission media characteristics (resolution, displaying distance, extend, color coding, ...). In fact, maps intended for distribution on the Web do hardly consider all different parameters for a successful customer perception. Instead “one map fits

all” paradigms are implemented. Selections with regards to content reduce a direct information load in maps, but do not effectively care about different use cases and perceptual parameters. One significant reason is that many use cases are not known in the map production process. Therefore use case identification and -consideration is needed within the map production flows (Farkas 2017).

Environments for map consumption as well as map production continuously change. Interaction paradigms (touch screens, virtual reality, ...) reach new dimensions. Potential digital map surfaces grow. For example the navigation-touch-screen of modern electrical cars has been extended to at least a sheet of A4 paper. Additionally latest developments of smartphones confront the dissemination of digital maps with foldable displays (Tan 2015). On the other side of map production, the digital spatial consumption counts on spatial swarm intelligence (Zhao 2016), an immediate and situation aware spatial processing within living information networks on the basis of spatial core datasets, that have been prepared and adopted for geo-communication.

The aforementioned confrontations in map production lead to considerations of

- handling logistical dependencies in knowledge networks of spatial swarm intelligence,
- making use of natural processing languages for the automation of information preparation and data wrangling and
- alternative methods for embedding resilient and reliant IT infrastructures.

2.1 Map production networks with logistical dependencies

Most of the activities of modern map production relate to FAIR data, decentralized data sources, data integration, service-oriented architectures, semantic collaboration, commonly usable knowledge networks, design principles and impact on user's behaviour and knowledge. This means that "map production" is developing towards production-process networks with logistical dependencies.

Losing sight of logistical dependencies within map production or within spatial knowledge networks leads to massive shortages. These shortages concern data quality- or service quality inconsistencies. For example if parts of data deliveries are too slow and therefore cannot be embedded in the map production process, these missing data will result in inconsistencies and malfunctions of the map: routing may not work, addresses do not show up,

one-way directions are wrong and so on. On the other hand, most cases of service quality inconsistencies are directly observed at the user: the content of services does not show up in the map because of bandwidth shortages. Therefore it makes the map product unusable.

Map production can overcome these shortages by embedding methods of supply chain. A well considered pool of data caches and logistical constraints leads to more stable map production infrastructures that could even support use cases of artificial intelligence, like in near future automotive applications. Furthermore it allows planning component changes and partial infrastructure failures (Wang 2016).

2.2 Natural Language Processing serving map production

Map compilation/production atop spatial data infrastructures requires comprehensive and mature understanding of cross-thematic data sources, their qualities, processing abilities, requested communication aim and transmitting device characteristics. Different communities and/or domains will name the same “thing” different. For example stakeholders of the environmental spatial data infrastructure in Europe will speak of a “view service”, whenever they mean an OGC web map service. Different names for almost the same meaning are common in decentralized infrastructures. In the end these differences cause inconsistencies within the network or even data. For some issues, like data identification and information preparation/wrangling, Natural Language Processing (Collobert 2011) could help to automatise parts of the production chain.

Especially the parsing and (geo-)coding functionality of Natural Language Processing can help for information retrieval that is applied on unstructured text, but based on structured analysis criteria (Purves et al 2005).

2.3 Decentralized networks and their required IT infrastructure

The map production with distributed sources, real-time data collection and in a highly automated way can be described as Service-Oriented Mapping. Service-Oriented Mapping offers the creation of maps and complex geoinformation analysis by use of the Service-Oriented Architecture paradigm, which covers specific information technology characteristics (Bieberstein 2008) in a decentralized network. Decentralized networks, which may also contain centralized components, help to access more diverse data sources, support flexible production flows and enhance storage needs as well as processing

power, if appropriate standards, interfaces and connections are established.

Spatial data infrastructures that play their role as secure and reliable entry point to spatial data and -service nodes ask for continuous effort in their IT architecture. Often the fundament of a spatial data infrastructure, the IT architecture covering security-, functionality-, interface-, processing- and provision tasks is taken for granted. But the main costs for modern map production and digital maps are generated in this back-office tier. When it comes to service-oriented map production, the IT architecture, its composition and flexibility decides about the success of the map.

3. Geoinformation processing

Due to growing data collections, diversity of data sources and an increasing number of use cases, any automation of geoinformation processing is needed. It follows aspects of data integration, data interpretation (Sester 2010), data “semantification” (Mami et al 2016) and data transmission preparation. For the variety of components, that provide functionality, Service-Oriented Architecture (SOA) is the main paradigm to connect and explore everything (Hendriks 2012). A central characteristic of the Service-Oriented Architecture is its loose coupling of components which allows component flexibility. This characteristic is made possible with standardized interfaces which connect single components.

Data semantification adds multilingual- and contextual meaning as well as importance to data. Any information processing with the aim of data semantification supports linguistic independence and enhances contextual data integration.

In addition geoinformation processing supports the metamorphosis from data- to knowledge transmission in terms of efficient communication for all embedded information handlers.

3.1 Data integration

Data integration is an urgent need: Different thematic communities make use of maps, which highlight the importance of involved geoinformation. The spatial dimension is mainly used to structure information by space, to show spatial dimensions or to make it comparable in time, respectively for multiple dimensions, like it is done in datacubes (Baumann 2018). This enrichment of thematic information by its spatial dimension could be called "geocoding" or "georeferencing", which enables additional integration,

comparison of data and knowledge extraction by its location.

“Data integration” or "data combination" is understood as the merging, processing and creation of datasets (UNGGIM 2018), which do not restrict sources, characteristics, qualities, extensions or specifications. From this wide-open point of view, data combination can be done from various sources. Even data aggregation of the same dataset with itself at different quality levels will deliver a new combined dataset. It incorporates lists, tables, geometric simple features and complex thematic objects and therefore goes beyond simple data table joining.

3.2 Discretized information processing

Discretization of spatial data is gaining importance in geoinformation processing and for data integration. The reason is the demand for direct data comparison in n-dimensions, at least for the dimension time (Evangelista 2018). Domain-independent and comparable grid systems are core for any information comparison. It does not matter, if information is delivered by the statistical, geological or meteorological domain. Although different histories resulted in diverging grid system definitions, nowadays demand urgently calls for harmonisation of the existing grids.

Datacubes extend dimensional restricted grids as a means to facilitate analysis and management of very large volumes of structured spatial data. With a global agreed and harmonised definition of “datacube”, collaborations that exchange function instead of data become feasible. Therefore the datacube concept has been deconstructed into distinct aspects, the six independent and orthogonal faces of a datacube (Strobl et al 2017).

The drawback of orthogonal grid definitions is their restriction for polar zones, which require appropriate grid definitions that do not match with the continental ones. Global grid systems seem to be a solution because they use a hierarchical tessellation of cells to partition and address the entire globe. Hexagonal discrete global grid systems provide an organisational structure for integrating multiple sources of large and variable geospatial information. This integration is sufficient for visualization and analysis (Purrs et al 2016, Ben et al 2018).

4. Enhancing Service-Oriented Map Production with the Semantic Web

The topic of digitalization is characterised by automatized procedures, precise information extraction, wide-spread relations of integrated knowledge networks and adaptive reasoning. Knowledge networks have been developed in the domain of artificial intelligence, where the description of knowledge in a machine-processable way is investigated. These efforts of extending the Web with semantic processing shall also extend spatial data infrastructures in order to find, extract and link spatial content more efficiently. The W3C defines the semantic web as "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries". One obvious requirement of these scientific networks is the mutual understanding of all embedded information. This calls for easy accessible and automatic processable knowledge networks as a pragmatic start of the spatially related semantic web (Egenhofer 2002).

4.1 Artificial intelligence and SOA

The geospatial fundament for automatization and artificial intelligence is a service-oriented architecture, in which no human decisions are needed for selecting the right content. Instead metadata, codelists, registries, ontologies and graphs deliver basic brickstones to establish a network for machine learning and artificial geospatial intelligence. This technological framework requires "new" storage methods as well as expanded representations of geographies, which have to be delivered by map production networks. At least these expectations for open- findable-, accessible- and reusable spatial information govern additional directions of map production and geoinformation management.

4.2 Spatial knowledge creation

Automated geospatial intelligence seems to be an additional role for cartography, which covers spatial knowledge creation and transmission between machines as well as humans.

Geospatial artificial intelligence (geoAI) with its principal task of data analysis and extraction, combines innovations in spatial science and artificial intelligence methods. These methods cover deep learning, data mining as well as high-performance computing to extract information from spatial big data (VoPham et al 2018). GeoAI researches the ability to connect large amounts of big spatial and temporal data in a variety of standardized formats. It includes computational efficiency, flexibility

for algorithms and openness for workflows that have to accommodate emerging characteristics of spatial processes, and owns scalability to model other requirements across the geographic domain (IBM 2017).

The cartographic perspective of transmitting geospatial knowledge has slightly other requirements than those of analysis geographic processes: the content has to perceptively and cognitively reach the user. Therefore we could formulate these requirements in the field of map production and create stable production processes that make use of geoAI.

5. Conclusion

We can clearly conclude that map production is confronted with emerging paradigms, methods and technologies. The paradigms make use of decentralized networks, which assume persistent IT infrastructures and interconnectedness. Methods in natural language processing could help to overcome different descriptions for the same meaning and therefore establish spatial semantification. Technologies concern decentralized map production networks in which logistical dependencies occur. Considering supply chain methodologies in combination with caching mechanisms to solve logistical dependencies could create a more efficient and fluid user experience for digital maps.

GeoAI develops as driver for modern map production. Although the intention for this discipline is information extraction and connecting brick stones of knowledge, map production could make use of geoAI approaches and procedures. At least cartography will have to transmit newly discovered knowledge with a smart selection of graphical variables and design.

The aim of this contribution and starting discussion is to formally extend the function of cartography by the aspect of artificial knowledge expansion and therefore to highlight the basics, requirements and additional emerging methods in map production.

6. References

- Baumann, P. (2018). Datacube Standards and Their Contribution to Analysis-Ready Data. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 2051-2053). IEEE.
- Ben, J., Li, Y., Zhou, C., Wang, R., & Du, L. (2018). Algebraic encoding scheme for aperture 3 hexagonal discrete global grid system. *Science China Earth Sciences*, 61(2), 215-227.

- Bieberstein, Norbert and Laird, Robert and Jones, Keith and Mitra, Tilak (2008) Executing SOA: a practical guide for the service-oriented architect.
- Collobert, Ronan and Weston, Jason and Bottou, Leon and Karlen, Michael and Kavukcuoglu, Koray and Kuksa, Pavel (2011) Natural language processing (almost) from scratch.
- Egenhofer, Max J (2002) Toward the semantic geospatial web.
- Evangelista, P. F., & Beskow, D. (2018). Geospatial Point Density. *R Journal*, 10(2).
- Farkas, G. (2017). Applicability of open-source web mapping libraries for building massive Web GIS clients. *Journal of Geographical Systems*, 19(3), 273-295.
- Hendriks, Paul HJ and Dessers, Ezra and Van Hootegem, Geert (2012) Reconsidering the definition of a spatial data infrastructure.
- IBM. Industry Insights: 2.5 quintillion bytes of data created every day. How does CPG & Retail manage it? <https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/>. Accessed 30 Oct 2017.
- Juckles, M., Swart, R., Barring, L., Groot, A., Thyse, P., Som de Cerff, W., ... & Bennett, V. (2016). A Climate Information Platform for Copernicus (CLIPC): managing the data flood. In EGU General Assembly Conference Abstracts (Vol. 18).
- Mami M.N., Scerri S., Auer S., Vidal ME. (2016) Towards Semantification of Big Data Technology. In: Madria S., Hara T. (eds) Big Data Analytics and Knowledge Discovery. DaWaK 2016. Lecture Notes in Computer Science, vol 9829. Springer, Cham
- Natural Language Processing in Geographic Information Systems – Some Trends and Open Issues –. Available from: https://www.researchgate.net/publication/232250058_Natural_Language_Processing_in_Geographic_Information_Systems_-_Some_Trends_and_Open_Issues [accessed Apr 09 2019].
- Purss, M. B., Gibb, R., Samavati, F., Peterson, P., & Ben, J. (2016). The OGC® Discrete Global Grid System core standard: A framework for rapid geospatial integration. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3610-3613). IEEE.
- Purves, R.; Clough, P.; Joho H. (2005) Identifying imprecise regions for geographic information retrieval using the web, in R. Billen, J. Drummond, D. Forrest, E.M João, and University of Glasgow, editors, Proceedings of the GIS Research UK 13th annual conference GISRUK 2005, Glasgow, Schottland, April 06-08,2005, pages 313–318,University of Glasgow, 2005.
- Sester, Monika (2010) Geoinformatics.
- Strobl, P., Baumann, P., Lewis, A., Szantoi, Z., Killough, B., Purss, M. & Dhu, T. (2017). The Six Faces of The Datacube. In Proc. Conf. on Big Data from Space (BiDS'17) (pp. 28-30).
- Tan, D., Kumorek, M., Garcia, A. A., Mooney, A., & Bekoe, D. (2015). Projectagami: A foldable mobile device with shape interactive applications. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1555-1560). ACM.
- UNGGIM-Europe (2018) UNGGIM Europe WG B3: Report on side-effects induced by data combination.
- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y. Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*, 17(1), 40.
- Wang, J., Muddada, R. R., Wang, H., Ding, J., Lin, Y., Liu, C., & Zhang, W. (2016). Toward a resilient holistic supply chain network system: Concept, review and future direction. *IEEE Systems Journal*, 10(2), 410-421.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3.
- Zhao, X., Lin, W., Hao, J., Zuo, X., & Yuan, J. (2016). Clustering and pattern search for enhancing particle swarm optimization with Euclidean spatial neighborhood search. *Neurocomputing*, 171, 966-981.