Abstract Data Types for Spatio-Temporal Remote Sensing Analysis

Martin Sudmanns
University of Salzburg, Department of Geoinformatics - Z_GIS, Schillerstraße 30, Salzburg, Austria
martin.sudmanns@sbg.ac.at
https://orcid.org/0000-0002-0473-1260

Stefan Lang
University of Salzburg, Department of Geoinformatics - Z_GIS, Schillerstraße 30, Salzburg, Austria
https://orcid.org/0000-0003-0619-0098

Dirk Tiede
University of Salzburg, Department of Geoinformatics - Z_GIS, Schillerstraße 30, Salzburg, Austria
https://orcid.org/0000-0002-5473-3344

Christian Werner
University of Salzburg, Department of Geoinformatics - Z_GIS, Schillerstraße 30, Salzburg, Austria

Hannah Augustin
University of Salzburg, Department of Geoinformatics - Z_GIS, Schillerstraße 30, Salzburg, Austria

Andrea Baraldi
Italian Space Agency (ASI), Rome, Italy.

Abstract

Abstract data types are a helpful framework to formalise analyses and make them more transparent, reproducible and comprehensible. We are revisiting an approach based on the space, time and theme dimensions of remotely sensed data, and extending it with a more differentiated understanding of space-time representations. In contrast to existing approaches and implementations that consider only fixed spatial units (e.g. pixels), our approach allows investigations of the spatial units' spatio-temporal characteristics, such as the size and shape of their geometry, and their relationships. Five different abstract data types are identified to describe geographical phenomenon, either directly or in combination: coverage, time series, trajectory, composition and evolution.

2012 ACM Subject Classification Information systems → Search interfaces

Keywords and phrases Big Earth Data, Semantic Analysis, Data Cube

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.60

Category Short Paper

Funding The study was supported by the Austrian Science Fund (FWF) through the Doctoral College GIScience [DK W1237-N23] and by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) under the program "ICT of the Future" within the project SemEO [contract no: 855467].

10th International Conference on Geographic Information Science (GIScience 2018).
Editors: Stephan Winter, Amy Griffin, and Monika Sester; Article No. 60; pp. 60:1–60:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction & Motivation

In the context of big Earth data, users do not seem to struggle mainly with technical problems, such as the provision of hardware (e.g. disk space or computing power), but are challenged by conceptual problems. These include decisions on how to observe phenomena on Earth (e.g. see [6]), store and analyse observations (e.g. see [3]), or replicate studies (e.g., see [19] or [14]). The value of big data, other than their volume, variety, and velocity, is challenging to leverage not based on inherent data characteristics, rather by how the data will be used [13]. For example, many data storage systems perform well when inputting data (i.e. saving raw EO images), but perform poorly when outputting data (i.e. finding relevant data and producing information from them) [13, 22]. Not knowing how data are structured and how they will be used on a generic level does not only challenge the general use of big Earth data, but also the replication of studies and reuse of workflows, because tools are not clearly distinguished from methods and data are not separated by semantic type [19].

Regular, free provision of Landsat and Sentinel data makes analyses of the temporal dimension increasingly important. Therefore, 3D Earth observation (EO) geospatial data cubes [18, 17] are becoming an increasingly popular tool. They do not treat images as temporally isolated, but index and reference them in a data structure where all axes (e.g. spatial and temporal dimensions) can be integrated and accessed equally [18]. It is necessary to know what types of queries are expected in order to decide on an optimal tiling scheme to optimize a geospatial data cube [8].

Increased data availability allows for analysis of high-resolution images, like Sentinel-2, on a continental or global scale, therefore opening new application domains such as serving the information needs of intergovernmental agreements, e.g. the United Nations sustainable development goals (SDGs). In this context, EO data and analysis methods spread into 'new' domains and confront new user communities with their complexity and particularities without providing a guiding and logical understanding of the representation of the geographical reality.

With all the technical preconditions available, analyses still aim to produce information relevant to questions posed by humans. The translation from questions to queries and results to answers is difficult, necessitates more than increasing data volumes and computing power, and goes beyond pure technical achievements. Recent developments are often technology-driven and are not necessarily tied to user requirements, where user groups are also non-experts from various application domains. For example, terminologies like 'big Earth data', 'data cube' or 'analysis ready data' are used before a proper definition or a common understanding is achieved. Inexperienced users struggle to become familiar with tools for reasons which might include a lack of common core terminology [15] and gaps between the user domain and the technical EO image domain [22, 21, 4]. This is especially complicated because a consistent conceptual model of space-time (e.g. consisting of continuants and occurrents (events) [9] and their relationships), as a representation of a mental model of the physical world (i.e. world model or world ontology), is still missing.

While the definition and a formalisation of a world model goes beyond the scope of this short paper, a certain level of understanding of at least continuants is necessary as a first step. A continuant can be seen as an entity in the physical world, parameterised by a unique continuant-identifier and an inner state, consisting of three types of attributes in the modelled 4D physical world: (a) positional, 3D geospatial attributes in geospatial units (e.g. lat-long coordinates and height in meters); (b) time attribute in a physical unit of time; and (c) "theme" [20]. We define theme as the combination of: (I) a theme type (i.e. geo-objects, geo-fields, and field-objects according to [10]); (II) a theme name (e.g.

any symbolic geo-object has a theme name belonging to a finite and discrete hierarchical, structured taxonomy of concepts or classes of real-world objects); and (III) appearance properties in the 4D physical world, expressed as either quantitative/numeric variables or qualitative/categorical sub-symbolic theme attributes in physical units [16]. These are: (1) photometric properties, expressed as either numeric colour values in spectral reflectance units (e.g. mean reflectance) or categorical colour names (e.g. red) belonging to a community-agreed discrete and finite vocabulary of colour names, related to a partition of a numeric colour space into quantization bins [11]); (2) shape (i.e. geometric) variables [2] such as compactness, rectangularity, elongatedness, straightness of boundaries, simple connectivity and orientation; and (3) size variables, like length and width in metres. Occurrents, as events, are able to change the inner state of a continuant, its relationship to other continuants, or the emergence of new continuants. To stick with the examples given above, we may conceive occurents as rotating crop types on an agricultural field, or the vanishing of a lake. The latter changes its size and thereby also its relationship to other continuants (patches of vegetation or open soil), which emerge simultaneously as new continuants.

For defining abstract data types for the application on Earth observation data, our conjecture is:

1. *The variety of phenomena in the focus of Earth observation can be represented and categorised by a limited set of abstract data types.*
2. *Having a set of defined abstract data types and knowing their behaviour can make remote sensing analyses more comprehensive and reproducible.*

2 State-of-the-art and research gap

A set of generic data types for spatio-temporal data was proposed by [7] based on three dimensions (i.e. spatial, temporal and thematic dimensions) inherent to any geospatial data [20]. Observations can be analysed by keeping one attribute fixed, controlling another and measuring the third. For example, in an EO image, fixing time, but controlling space and measuring the theme yields a land cover map. Similarly, fixing space (e.g. the location of a temperature sensor), controlling time and measuring the theme represents a temperature curve throughout a year. In total, [7] identified three out of nine possible data types as relevant:

- Coverage: fixing time, controlling space, measuring theme
- Time series: fixing space, controlling time, measuring theme
- Trajectory: fixing theme, controlling time, measuring space

Another method for separating geospatial data types from their physical organisation is comprehensively described by [1], where "spatial lenses" provide software-based views as a way to interpret datasets. The interpretations, based on a specific view of the world, include a network, objects, fields and events, as well as refer back to the core concepts of spatial data [15].

In the remote sensing domain, geographic object-oriented image analysis (GEOBIA) uses image segments (i.e. objects) instead of pixels as target analysis units [5]. Therefore, GEOBIA applies object-oriented data models to geographic image data. Since the segments have inherent spatial characteristics (e.g. size, shape, topological arrangement) and can be temporally associated with each other, GEOBIA allows spatial and temporal analyses. Typically, the objects' semantics are modelled using ontologies or a rule-based approach, such as implemented in the eCognition software. However, the ontologies or rule-sets are usually tied to a virtual 2D map legend domain and not to the 4D physical world domain [4].

Separating the virtual image domain from the physical world domain in EO image analysis was introduced in [16] and was then later taken up and applied as a GEOBIA-based approach by [12] and [22, 21, 4].

Although some previous work is available, a set of universally applicable, comprehensive, abstract data types for EO data have not yet been developed. Such a set could serve as a framework for mapping spatial, temporal and thematic attributes of observations in EO data cubes. Existing approaches and implementations lack either generality (e.g. specific GEOBIA implementations), or are limited to fixed analysis units (e.g. pixels). We suggest abstract data types to be used as a logical, intermediate layer between EO data cubes and the 4D physical world domain, thus adopting a clear distinction from the physical organisation of data [1] as well as the 2D virtual image domain [22, 21, 4]. Our proposed abstract data types adapt the ideas of [7] and extend them with the more differentiated understanding of space-time phenomena and their spatial, temporal or semantic relations in GEOBIA required for spatial image analysis [2]. Space in an EO image context has multiple meanings since it: (1) refers to the absolute or relative location of an object (e.g. represented by a coordinate tuple) and its spatial relation to other objects; and (2) also refers to inherent spatial characteristics of an object (e.g. size and shape). In a more complex situation, e.g observing the expansion of a city, the object itself is the result of a spatial arrangement of other objects, including houses and streets.

3 Proposed abstract data types

We differentiate between *position* (or *location*) and *space*, which are inherent spatial properties of objects. Further, a position of an object might not only be the absolute position, but also its relative location within a topological arrangement. We also differentiate between continuous (i.e. quantitative) and discrete (i.e. categorical) variables. The temporal dimension has its upper limit in t_0 and goes back until t_{-x} as this approach is intended for querying an archive and not for projecting processes in the future. The following abstract data types can be selected, and are illustrated in Figure 1:

- **Coverage**: constructed by fixing time, controlling position, measuring theme (continuous or discrete)
- **Composition**: constructed by fixing time, controlling theme, measuring space
- **Time series**: constructed by fixing the position, controlling time, measuring theme (continuous or discrete)
- **Trajectory**: constructed by fixing theme, controlling time, measuring position
- **Evolution**: constructed by fixing theme, controlling time, measuring space.

4 Conclusion & Outlook

Challenges of big Earth data go beyond technical issues. We suggest a limited, yet defined and tangible set of abstract data types, which are specifically selected for use as a framework for query primitives within EO data cubes. While existing solutions rely on fixed spatial units, such as pixels, in GEOBIA the space properties do not only refer to the position, but also to the spatial arrangement of objects and to properties such as extent, shape and size of the object under consideration. Based on the state-of-the art review, we found the necessity to extend the original set of abstract data types with two new ones to account for the differentiated view on space within the GEOBIA domain. While defining this framework is an ongoing process and this contribution is a first step towards it, in this short paper

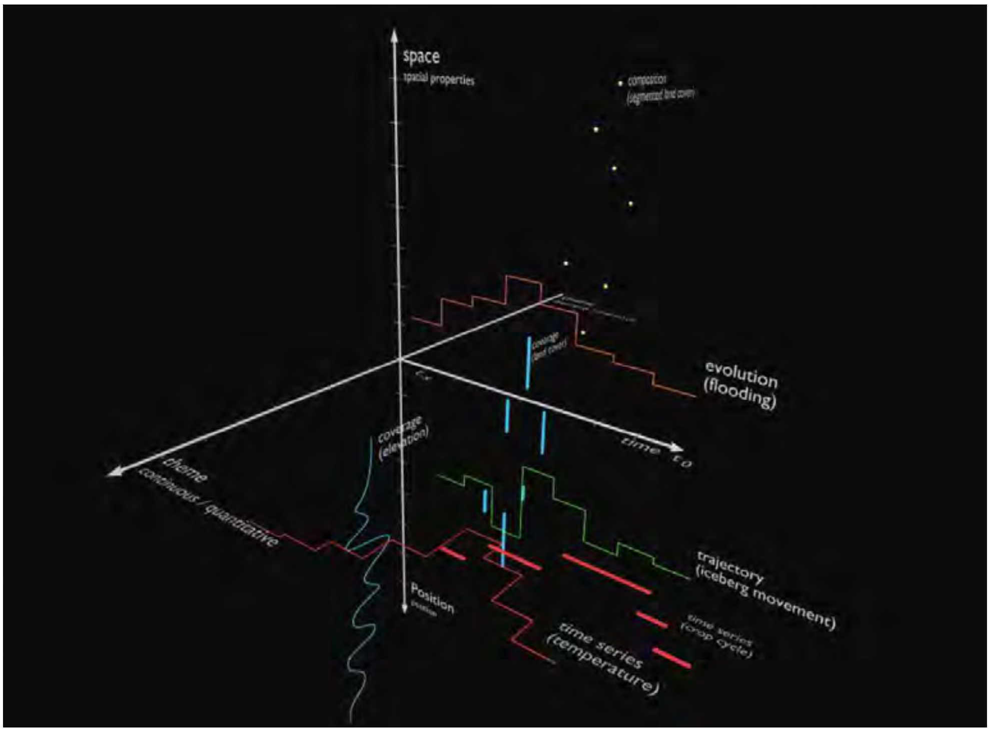

Figure 1 Example phenomena relevant to Earth observation visualised in a 3D space. Here, the axes provide an ordering principle for EO spatio-temporal phenomena. Note that space can be conceived as position (e.g. 0-dimensional, coordinate tupel or tripel) and the spatial relationship to other entities, or as geometric features (e.g. set of coordinate tupels, size, compactness). Although the attributes are represented on single, individual axis, the semantics of the axes differ between theme or time (monodimensional) and space or position (multidimensional). An interactive visualisation is available as online visualisation (http://cf000008.geo.sbg.ac.at/adt/).

we aim to highlight the necessity of having it for formalising queries. Future work will align this framework with the definition of a world model as a conceptual description of geospatial phenomena, e.g. using a rigorous formalisation of continuants, occurrences and their relationships. Further, this also includes revisiting the original and suggested terms and a discussion of whether they are appropriate for this purpose. Being in a preliminary stage, the framework and the abstract data types are presented here in a rather informal manner. Therefore, the focus will lie on the formalisation of the data types and their methods as well as an example implementation in an EO data cube.

Abstract data types allow for semantic annotations and workflow exchanges by separating methods from tools and the image domain from the physical world domain. They can be considered as a logical, intermediate layer between the conceptual world model and the data storage engine, e.g. geospatial data cubes. Therefore, they can be used to answer questions such as “what data are used?" or "what are they useful for?” and are linked to big Earth data relevant decisions. These include but are not limited to how certain phenomena can be observed, how a system can be designed to provide analysis results with reasonable response times and how the result can be interpreted and deemed trustworthy. Further, they help non-EO experts to express their questions in formalised terms. It is increasingly relevant to analyse EO data together with non-EO data, where abstract data types might also play a significant role.

References

1 Christopher Allen, Thomas Hervey, Sara Lafia, Daniel W. Philips, Behzad Vahedi, and Werner Kuhn. Exploring the notion of spatial lenses. In *Geographic Information Science 2016*, pages 260–274. Springer, Montréal, Canada, 2016. doi:10.1007/978-3-319-45738-3.

2 Andrea Baraldi, Dirk Tiede, Martin Sudmanns, and Stefan Lang. Systematic esa eo level 2 product generation as pre-condition to semantic content-based image retrieval and information/knowledge discovery in eo image databases. In Publications Office of the European Union, editor, *Proceedings of the 2017 conference on Big Data from Space*. Publications Office of the European Union, 2017.

3 Peter Baumann, Paolo Mazzetti, Joachim Ungar, Roberto Barbera, Damiano Barboni, Alan Beccati, Lorenzo Bigagli, Enrico Boldrini, Riccardo Bruno, Antonio Calanducci, Piero Campalani, Oliver Clements, Alex Dumitru, Mike Grant, Pasquale Herzig, George Kakaletris, John Laxton, Panagiota Koltsida, Kinga Lipskoch, Alireza Rezaei Mahdiraji, Simone Mantovani, Vlad Merticariu, Antonio Messina, Dimitar Misev, Stefano Natali, Stefano Nativi, Jelmer Oosthoek, Marco Pappalardo, James Passmore, Angelo Pio Rossi, Francesco Rundo, Marcus Sen, Vittorio Sorbera, Don Sullivan, Mario Torrisi, Leonardo Trovato, Maria Grazia Veratelli, and Sebastian Wagner. Big data analytics for earth sciences: the earthserver approach. *International Journal of Digital Earth*, 9(1):3–29, 2016. doi:10.1080/17538947.2014.1003106.

4 Mariana Belgiu, Martin Sudmanns, Dirk Tiede, Andrea Baraldi, and Stefan Lang. Spatiotemporal enabled content-based image retrieval. In *Ninth International Conference on GIScience, Short Paper Proceedings*, volume 9. University of California, 2016.

5 Thomas Blaschke, Geoffrey J Hay, Maggi Kelly, Stefan Lang, Peter Hofmann, Elisabeth Addink, Raul Queiroz Feitosa, Freek van der Meer, Harald van der Werff, Frieke van Coillie, et al. Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191, 2014.

6 M Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sensing of Environment*, 120:25–36, 2012.

7 Karine Reis Ferreira, Gilberto Camara, and Antônio Miguel Vieira Monteiro. An algebra for spatiotemporal data: From observations to events. *Transactions in GIS*, 18(2):253–269, 2014.

8 Paula Furtado and Peter Baumann. Storage of multidimensional arrays based on arbitrary tiling. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 480–489. IEEE, 1999.

9 Antony Galton. Experience and history: Processes and their relation to events. *Journal of Logic and Computation*, 18(3):323–340, 2008.

10 Michael F Goodchild, May Yuan, and Thomas J Cova. Towards a general theory of geographic representation in gis. *International journal of geographical information science*, 21(3):239–260, 2007.

11 Lewis D Griffin. Optimality of the basic colour categories for classification. *Journal of the Royal Society Interface*, 3(6):71–85, 2006.

12 Stefan Growe. Knowledge based interpretation of multisensor and multitemporal remote sensing images. *Int. Arch. Photogramm. Remote Sens*, 32(pt 7):4–3, 1999.

13 Adam Jacobs. The pathologies of big data. *Communications of the ACM*, 52(8):36, 2009. doi:10.1145/1536616.1536632.

14 Christian Knoth and Daniel Nüst. Reproducibility and practical adoption of geobia with open-source software in docker containers. *Remote Sensing*, 9(3):290, 2017.

15 Werner Kuhn. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science*, 26(12):2267–2276, 2012. doi:10.1080/13658816.2012.722637.

16 T. Matsuyama and V.S.S. Hwang. *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Advances in Computer Vision and Machine Intelligence. Springer US, 1990.

17 Stefano Nativi, Paolo Mazzetti, and Max Craglia. A view-based model of data-cube to support big earth data systems interoperability. *Big Earth Data*, pages 1–25, 2017.

18 Peter Strobl, Peter Baumann, Adam Lewis, Zoltan Szantoi, Brian Killough, Matthew Purss, Max Craglia, Stefano Nativi, Alex Held, Trevor Dhu. The six faces of the data cube. In Publications Office of the European Union, editor, *Proceedings of the 2017 conference on Big Data from Space*, pages 32–35. Publications Office of the European Union, 2017.

19 Simon Scheider, Frank O. Ostermann, and Benjamin Adams. Why good data analysts need to be critical synthesists. determining the role of semantics in data analysis. *Future Generation Computer Systems*, 72:11–22, 2017. doi:10.1016/j.future.2017.02.046.

20 David Sinton. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. *Harvard papers on geographic information systems*, 6:1–17, 1978.

21 Martin Sudmanns, Dirk Tiede, Stefan Lang, and Andrea Baraldi. Semantic and syntactic interoperability in online processing of big earth observation data. *International Journal of Digital Earth*, 11(1):95–112, 2018. doi:10.1080/17538947.2017.1332112.

22 Dirk Tiede, Andrea Baraldi, Martin Sudmanns, Mariana Belgiu, and Stefan Lang. Architecture and prototypical implementation of a semantic querying system for big earth observation image bases. *European journal of remote sensing*, 50(1):452–463, 2017. doi:10.1080/22797254.2017.1357432.