The Semantic Pixel

Fred Fonseca[1] Clodoveu Davis[2] Gilberto Câmara[3]

[1]College of Information Sciences and Technology – The Pennsylvania State University – 307E IST Building – University Park, PA – USA

[2] Departamento de Ciência da Computação – Universidade Federal de Minas Gerais Av. Antônio Carlos 6627 - Belo Horizonte - Brazil

[3] INPE - Brazilian National Institute for Space Research Av dos Astronautas, 1.758 – São José dos Campos – SP - Brazil

fredfonseca@ist.psu.edu, clodoveu@dcc.ufmg.br, gilberto.camara@inpe.br

Abstract. *Usually, images can be seen as sets of pixels or as fields over a reference space. While the former view allows image processing to function using pixel manipulation algorithms, the second one is closer to a wider understanding of what people perceive in an image. The pixel aspect is much closer to the measurement, to observations, while fields are closer to the semantic aspect, to the interpretation of the observations. This paper discusses some semantic challenges related to integration of image data from various sources, considering both views. Such integration is necessary, considering that soon a new generation of remote sensing satellites based on free and open data policies is expected to become operational, so researchers will have access to more data than they can handle with current techniques. We propose the integration of images from multiple sensors starting from a common point, which we call the Semantic Pixel. It will enable scientists to have access to large sets of satellite images and their metadata, regardless of source or format. The Semantic Pixel will also enable access to ancillary data, which is essential for advanced temporal analysis of forest cover dynamics, including major sets of natural resource data, such as vegetation, soil and geology maps. Other data encoded as fields, such as digital elevation models, relevant climatic variable maps, political maps and associated census data, can also fit this model.*

Resumo. *Imagens podem ser vistas como um conjunto de pixels ou como campos em um espaço de referencia. Enquanto os pixels permitem que algoritmos de processamento de imagens possam funcionar, os campos estão mais próximos do que as pessoas entendam o que seja o significado de uma imagem. A visão de pixels está muito mais próximo das medidas, das observações, enquanto os campos estão mais próximos da semântica, da interpretação das observações. Este artigo usa estas duas visões para discutir os desafios relativos a semântica na questão das integração de imagens de provenientes de fontes diversas. Esta integração é necessária já que brevemente novos satélites com politicas de dados abertos devem estar disponíveis o que levará os cientistas a terem mais dados do que eles possam efetivamente usar. Aqui nós apresentamos uma proposta de integração de imagens de fontes diversas usando como plataforma inicial um ponto comum,* *que chamamos o Pixel Semântico. Desta maneira os cientistas teriam acesso a dados e metadados de imagens independente da fonte ou formato. O Pixel Semântico também possibilitaria o acesso a dados históricos que são importantes para análises da dinâmica da vida das florestas tropicais.*

1. Introduction

There is a consensus around the idea that the data needs for deforestation monitoring are so broad that the right way to approach this problem is to facilitate access to whatever data is available. There are certainly large amounts of valuable data collected for many scientific models on deforestation which are currently inaccessible, or worse, whose existence is practically unknown for potential users. In order to provide an adequate compromise between resolution and coverage as needed for forest assessment, a global forest monitoring system (Fonseca, Davis Jr. et al. 2009) would have to integrate satellite images of different kinds. For instance, a combination of up-to-date MODIS-class (250-meter), LANDSAT-class (30-meter) and HRC-class (2.5-meter or better) images would be required for many applications. Besides that, by 2015, a new generation of remote sensing satellites (LANDSAT-8, CBERS-3, Sentinel-2) is expected to become operational, based on free and open data policies. "The age of big geospatial data has come. Space agencies worldwide plan to launch around 260 Earth observation satellites over the next 15 years."(Ferreira, Câmara et al. 2014)

The motivation of this paper is to address this increasing volume of data from multiple and diverse satellites and support its use by scientists. We take a different stance from Werner Kuhn, who tries to model "the relevant information processes independently of sensor technology" (Kuhn 2009). We acknowledge the importance of Kuhn's approach but we also want to recognize the push from technology and measurement technologies, which seem to be driving the process of data collection.

We also want to address the new possibilities brought up by new and advanced database technologies, such as NoSQL array databases (Paul; Stonebraker 2010), that are enabling scientists to have large time series of remote sensing images available as a single multidimensional matrix. Conceptually, then, any remote sensing data can be referred to by matrix elements containing a value, a timestamp and a geometry description (Ferreira, Câmara et al. 2014). This way, any remote sensing data can be addressed and manipulated based on its most fundamental component (a pixel), instead of using full images or scenes. Pixel-wise data can be organized and handled with array databases such as SciDB[1].

Although the pixel is the basic unit for remote sensing images, the semantics lie at a much larger level. Themes, coverages, and relationships are all combinations of pixels across time, and combinations of pixels in themes and concepts across space. In this paper we focus on the challenges in starting from the basic unit of recording data, the pixel level, and going to the meaning of full observations represented by, for instance, time series, trajectory and coverages (Ferreira, Câmara et al. 2014). We want to know how pixel sets can be understood and become linked to higher semantic concepts. When the pure measurement-related pixel is regarded at a higher semantic

[1] http://www.scidb.org

level, it starts carrying more meaning than just a value, a timestamp and a location. It becomes the *semantic pixel*.

This paper is organized as follows. Section 2 reviews the relation between data and theories in science in order to understand the role that pixels play as raw data. Section 3 introduces the concept of the semantic pixel. Other aspects of the semantic pixel related to pure data, aggregation, and measurement are discussed in Section 4. Finally, Section 5 concludes the paper.

2. Data and Theories

In 1978, when GIS was starting, Sinton (1977) thought that the "largest groups of potential users will be those who already use geographic information in non-digital analytic formats". Today, Google Maps and many other geographic information sources are ubiquitous. Every phone has a GPS and you can ask for directions just by talking with your handheld device. Technology has a prevalence that is concerning.

Sinton argued that geographic information systems had to cater to user needs. It was a noble goal, although system and users are still fighting. The huge availability of data, spanning all the regions of the world and now also long periods of time, at a very reasonable (sometimes neglectable) cost, is starting to show that, maybe, neither systems nor users have the upper hand, and the data won after all. So, there must be solutions for a situation such as the one Sinton observed:

> *"The digital encoding of mapped thematic data can significantly affect its applicability by these potential user groups. Experienced analysts have been using traditional organizational structures for information storage and retrieval. Mapped thematic data is universally stored in a map record format. The analytic sub-system of a geographic information system will be more easily used if it reproduces these traditional record keeping formats. Neither polygon nor "bit plane" data structures use a map record format."* (Sinton 1977)

We have now a situation in which analysts have to adapt to the format of the incoming data. Our proposal of a semantic pixel tries to address some of these issues.

2.1 The Changing Priority between Data and Theory

Scientists used to think that data were the source of theories. Only direct access to the data could save and keep Science free from metaphysics. Logical positivists claimed that propositions are reducible to elementary propositions. They tried to show that elementary propositions are exclusively concerned with picturing empirical reality. They were in part motivated by Wittgensteins' work, that tried to show that empiricism could be founded on logic.

> *"The philosophy of the Viennese Circle is an empiricism established by logical methods. Briefly stated, it is established by showing that under analysis the meaning of concepts and propositions is, in every case, ultimately empirical. Propositions which are not ostensibly empirical in reference are therefore either reducible to empirical propositions or are simply nonsense".* (Weinberg 1960 p.25)

Later, philosophical discussions in the 20th century ended up showing that theories are created before data and therefore, theories drive the data collection process.

Now, with the increasing amount and availability of satellite data, is this scenario shifting in favor of data driving theory again?

We are not arguing for the importance of observations being driven by theory over technology driving the observation process. We are only acknowledging the deluge of data from satellites and proposing a way to incorporate it in our models of science. In a way, both data and theories contribute to Science's understanding of the world.

Again, we can use Werner Kuhn (Kuhn 2009) to clarify the problems with the current situation. While Kuhn considers observation to be "an information item with semantics that are independent of observation technology", we are explicitly acknowledging that today, with the large availability of current and historical satellite data, technology is playing a leading role in the generation of observations. It is a situation in which the technological focus is "putting encoding before modeling" (Kuhn 2009).

If we are really going back to data creating theories, this is an interesting change from the way Science works, usually with theory driving observations and not the other way around. Kant was the first to remind us that

> *"When approaching nature, reason must hold in one hand its principles, in terms of which alone concordant appearances can count as laws, and in the other hand, the experiments thought out in accordance with these principles. Thus reason must indeed approach nature in order to be instructed by it; yet it must do so not in the capacity of a pupil who lets the teacher tell him whatever the teacher wants, but in the capacity of an appointed judge who compels the witness to answer the questions that he puts to them"* (Kant 1996)

Later, Popper confirmed this approach in a famous footnote in The Logic of Scientific Discovery, telling us that we always gather facts with theories in mind. He says "… observations, and even more so observation statements and statements of experimental results, are always interpretations of the facts observed; that they are interpretations in the light of theories" (Popper 2002).

Thomas Kuhn (1996) and Paul Feyerabend (1993) also agree that there is, in general, no neutral set of facts that provide a framework for comparing competing theoretical perspectives. For these crucial areas of difference, the facts they would notice would not be neutral, but would have their existence only relative to a given paradigm. Thomas Kuhn insists that the notion of a deeper, and hence neutral, classificatory scheme, is illusory. All classifications of the facts are relevant to some theoretical framework, in Thomas Kuhn's view. A so-called deeper classificatory scheme would not be neutral but relative to its own theoretical perspective.

We might think that the accumulation of decades of remote sensing data and the increased availability of images might be changing the situation described above because now we have access to the pure facts again, and lots of them. But we have to remember that instruments are also supported by theories:

> *"Galileo claimed that he could 'observe' mountains on the moon and spots on the sun and that these 'observations' refuted the time honoured theory that celestial bodies are faultless crystal balls. But his 'observations' were not 'observational' in the sense of being observed by the-unaided-senses: their reliability depended on the reliability of his telescope-and of the optical theory*

of the telescope-which was violently questioned by his contemporaries. It was not Galileo's pure, untheoretical-observations that confronted Aristotelian theory but rather Galileo's 'observations' in the light of his optical theory that confronted the Aristotelians' 'observations' in the light of their theory of the heavens". (Lakatos 1970)

While hardware development and a wider satellite availability makes the pixel more complex one on side, new scientific theories and a deeper understanding of the Earth and its environment seems to create more complexity on the other side. The work in this paper is a step in the direction of understanding these two sides so that data enable better scientific theories and new scientific theories keep pushing the development of new measurement technologies.

In our case, even if remote sensing development might originate from pure Physics development, in a way unrelated to the sciences that will use the data, the understanding of the use of imagery data in Science is our goal. Therefore, data collection processes and the way such data are linked to theories are matters of interest. The discussion carried out in this Section is a step in understanding how satellite data at the pixel level can be kinked to complex concepts in scientific theories.

3. The Pixel as an Information Element

When trying to integrate images from various remote sensing sources, specialists are always confronted with problems regarding the compatibility of images as to spatial resolution, spectral resolution, and time. Putting together two images from different sources potentially requires the use of techniques and algorithms such as registration, radiometric correction, noise reduction, and others, usually applied in a pixel-by-pixel basis, and often causing distortions or information loss.

In the sense that an image is a representation of part of the world, however, concepts such as resolution should have no real value. Pixels should be viewed as an algorithm's way to break a larger problem (i.e., image processing) into many smaller ones (i.e., transforming pixel values in the vicinity of a point or inside a region). Pixels are samples to a phenomenon that extends well beyond their limits, and which cannot be perceived if we regard them individually.

As a result, there should be a level of semantic description of images in which pixels are irrelevant (Koenderink 2005). Operations on images could then be specified as functions that receive one or more representations of space and certain parameters, and generate other representations, regardless of how these representations are materialized in the first place. The most important concept in this situation is space and its characteristics; if we were able to represent every single point in space and associate it with measurements, the field representation (Couclelis 1992; Câmara, Freitas et al. 1994) would be perfect. Since we are not capable of doing so, we must deal with a coarser sampling, i.e, a measurable pixel size.

Problems begin to arise when one tries to integrate two of such representations, each of which obtained with a different set of parameters (sampling, resolution), with different sensors, at different times. In order to harmonize such differences, we propose the definition of an information element that is more generic than the pixels that compose current images, and that is strictly related to a small fragment of space. Each of these small fragments, which we call a *semantic pixel* (SP) from now on, is then associated to the various fields that are defined over its position, so that it is possible to find out information about that position from numerous sources. There should be only one SP for each fragment of space. All recorded phenomena that occur at that position are associated to the same SP, so that it is possible to select data and metadata at that position based on a set of requirements, such as a time interval, or the response to a given electromagnetic frequency range.

Each SP becomes, then, a generic reference to the many representations and measurements that are available for that position. Given a SP, it would be possible to recover the original images that include that position, as in an index, selecting among them from their metadata. In places where the information is missing or has been corrupted (for instance, when there is cloud cover over part of a RS image), algorithms or visualization could use data from other sources, also associated to the same SP. Examining SP data, users could select the attribute (or combination of attributes) which is more adequate to fulfill a given task. Furthermore, data from each SP used for integration can be traced back to its source, so that metadata apply to SP attributes, not only to an entire image. Results from combining and processing data from various sources at the SP would also be associated with provenance metadata, so that specialists could trace the outcome of analyses and processing to their origin.

One of the motivations for the creation of the semantic pixel is traceability. Instead of separately managing several images for one point in space, the SP is a representative of those images. The SP is only a link to the original and preserved images. It gives a common ground for visualization and navigation. This is especially important for remote sensing images, which can suffer from local (i.e., cloud cover, atmospheric effects) or global (georeferencing, registration) distortions, the correction of which always raises many questions.

A system using semantic pixels will have many images (raw and processed). The user has to have the final decision on what image to use in each algorithm. The SP will enable the user to do so. Different users may use the system in different ways. A scientist may want to use the raw satellite images, do all the pre-processing, segmentation and classification, apply her own algorithms and report her findings. Another scientist may use the pre-processed images already available and perform only classification and analysis using different algorithms for verification purposes. In both cases, the SP functions as both the starting point and the link between the region of study and all the available data.

The SP must also provide semantic interoperability. An application that intends to monitor deforestation needs high temporal resolution and adequate spectral resolution, but accepts low spatial resolution. On the other hand, an application that needs to measure deforestation requires a medium spatial resolution and adequate spectral resolution, and accepts a lower temporal resolution. Therefore, a deforestation monitoring pixel is different from a deforestation measurement pixel, since the sources of information are different. The SP has to provide for both cases.

Every SP corresponds to a geographic position, and is potentially related to many *content sets*. Each content set indicates one source of data for the position that corresponds to the SP. Content sets can be viewed as tables associated to each SP, in which each row corresponds to a data source and includes at least the following columns for data and metadata: source image URL, image timestamp, source image position, satellite (vehicle), sensor, band (frequency), spectral resolution (bits), and a Boolean indicator as to the position being cloud-covered. Value data, i.e., the measurements associated to the SP, are kept in the original image, regardless of being a raw image or a processed image (for instance, the results of a classification), but can be copied to the content set for faster retrieval. Further attributes can also be included, in order to maintain provenance data for processed images. Notice that, for every new image that is included in an archive, a new content set is created for each SP that is inside the image's spatial boundaries, but the image is also stored as received.

Looking at the resulting data infrastructure, notice that the set of SPs is highly redundant as to image metadata, in a strategy also successfully employed in tools such as data warehouses. As in data warehouses, we assume that computational resources to maintain such a level of redundancy are affordable enough to economically justify the expected semantic and performance gains. The objective is to allow the user to select a single SP, or a set of SPs, and retrieve the most adequate information available for that position or region, given a set of criteria, in a case-by-case basis. On the other hand, by grouping attributes such as the source URL together, it is possible to have direct access to the source of the information related to an SP.

4. The Many Dimensions of the Semantic Pixel

In this section we add other perspectives to the understanding of what the Semantic Pixel is. We discuss the operations that create the semantic pixel using Sinton's original ideas on the operations that pixels go through in order to be usable by scientists. Then we use Wittgenstein concept of object to discuss the semantic pixel as raw data and what it means in the scope of philosophy of science, in the middle of the discussion of theories and data. Finally we talk about the importance of understanding the pixel as measurement based on the Fonseca and Martin's (2007) conceptual framework of objects and objectives.

4.1 The Semantic Pixel is an Aggregation

Since the pixel, as most field observations are related to small sections of space and time, for its use in scientific analyses it is necessary some transformations. Sinton mentions the need for aggregation and also the problems that come with it:

> *"Original field observations tend to be voluminous in their nature. Consequently it is rare that so much data is reported and made generally available. The original collectors of the information will generalize it to reduce the volume and usually attempt to make it more understandable in an abstract form. The procedures used in the generalization or abstraction of data significantly affect its utility for analytic purposes."* (Sinton 1977)

A system such as a Global Forest Information System (Fonseca, Davis Jr. et al. 2009) using the semantic pixel would address two of the concerns Sinton had. First, regarding the user's understanding of the many transformations that raw data goes through before becoming available for scientific analyses. According to Sinton, "when presented with a set of data, individual users should attempt to understand the nature of the generalization that has taken place on the original observations or measurements" (Sinton 1977). In its original proposal, in a GFIS "there is the possibility of the different actors sharing and understanding the meaning of the scientific models explaining deforestation processes. This way information from the different sources can be used as a communication tool, in order to motivate common citizens, scientists, and the society at large, to contribute with the monitoring effort and to influence policy making and enforcement" (Fonseca, Davis Jr. et al. 2009).

GFIS would also bring aggregation to the user of the data again. With the new database techniques, users can have access to the pure pixel again, according to Gilberto Câmara, the Holy Grail of remote sensing. This way we can control the second concern mentioned by Sinton since "associated with each of these processes of generalization is the loss of certain types of detail which existed in the original observation or measurement". (Sinton 1977 p.4)

4.2 The Semantic Pixel is a Fact

In the discussions of Philosophy of Science, Wittgenstein was trying to understand the nature of analytic propositions and how they would fit with logics and at the same time be linked to the real world. For Wittgenstein,

> *"the world is the totality of independent atomic facts. An atomic fact is a fact which is not compounded out of other facts. Since facts are ultimately independent of one another all compound facts are reducible to atomic facts. Which facts are atomic and which are not cannot be determined a priori, but must, in any case, be discovered by direct inspection."* (Weinberg 1960 p.38)

It is difficult not to think of a similarity between "atomic facts" "discovered by direct inspection" and the pixels coming from satellite sensors. Are those pixels really the most basic facts about our world? Actually, another concept Wittgenstein developed, *object*, would be close to what a pure pixel is while *fact* should be related to what a Semantic Pixel is. For Wittgenstein,

> *An object is whatever can occur as the constituent of a fact. Now, if facts are taken as fundamental, and hence indefinable, an object could be variously defined, (a) It may be defined as the set of facts in which it occurs, i.e. as the set of facts which possess at least one feature of absolute similarity to one another. For example, the facts of blue colouring the sky at time t_0 and of blue colouring this book at time t_n have one feature, blue, of absolute similarity, (b) Or an object can be defined as whatever is a distinguishable element of a fact. Thus, by exhaustively enumerating all the distinguishable elements constituting a fact, it is possible to isolate all the objects composing the fact in question.*
>
> *The fact is an independent entity, for whatever dependence may mean in the strictly logical sense, it is reserved for objects, i.e. for entities obviously requiring completion. Facts, being self-sufficient, require no completion, and so are, in the logical sense, independent of one another. Objects are independent, too, in the sense that an object is not restricted to occurrence in one fact rather than another, but they are dependent in the sense that they must occur in some fact or other.* (Weinberg 1960 p.35-36)

Since the Semantic Pixel is one level away from the pure pixels, which are closer to Wittgensteinian objects, the SP it is better represented by facts. The Semantic Pixel, as a fact, is composed of objects (pure pixels).

4.3 The Semantic Pixel is Measurement

Although the pure pixel is the most basic unit for images, and it is meaningful in its own way, there is a big conceptual leap from the individual pure pixel to sets (aggregations) of pixels representing semantic concepts such as time series, trajectories or coverages.

The pure pixel's origin, more related to the measuring hardware than to its final semantic meaning, has to be understood. It has individual meaning and complexity before it becomes part of a larger semantic concept. The use of new database techniques such as array databases only reinforces this aspect. The pure pixel, coming from the measurement hardware, and the final aggregation of pixels in a semantic concept, belong to two different epistemic levels (Fonseca and Martin 2007). They are created with different objectives and they have different objects. The aggregated set deals with general assumptions concerning the explanatory invariants of a domain – those that provide a framework enabling understanding and explanation of data across all domains, inviting explanation and understanding. Aggregated sets belong to an ontological level (Fonseca and Martin 2007). Individual pure pixels are related to the consequent dimensions of possible variation among the relevant data of a given domain. There is a natural decomposition of information in terms of two necessary but complementary epistemic functions: identification of an invariant background (aggregated sets) and measurement of the object along dimensions of possible variation (the individual pure pixels).

Fonseca and Martin (2007) suggest the use of objectives and object to understand the difference between basic measurements and their later user as semantic concepts. Adapting their proposition to our case, we can say that regarding the *objectives*, aggregated sets provide explanation and information integration grounded in assumptions about invariant conditions that define the domain of interest. Pure pixels, on the other hand, enable the measurement and classification of the observed facts. Now, regarding the *object*, aggregated sets are based on the real world as understood by humans, on the reality, instead of focusing on what can be represented. The object is the representation of the invariant conditions of the domain of interest – the general, and assumed, categories that are taken to define a domain. Pure pixels are based on the permissible range of variation among the facts that, later, must be brought into relation with those categories.

5. Conclusions

For the Semantic Pixel to be used as a key to reach measurements, in the way of an index, it is necessary to define the kinds of operations that have to be implemented in order to achieve this kind of "transparency". In such a scenario, the user knows and mentions only the Semantic Pixel; then, a computational layer on the background translates the user's conceptualization into the values stored in the images and arrays. Initially, we think much in the way of the interpolation, resampling, segmentation and classification techniques that would be necessary. The difference is that the user would not need to know what is going on underneath the hood to get information adapted to his particular conceptualization. That can be defined as a large computational problem, since it must be efficient, automated, and would require large resources in terms of storage and processing. A good definition on what is required here could be put against what is available (and what can be proposed and implemented) as operations in SciDB, and would generate a to-do list for enhancing array databases to be used in large remote sensing libraries.

As we see it, the implementation of the Semantic Pixel could be something along the lines of the multidimensional arrays of SciDB (Paul; Stonebraker 2010), in which each cell has values in many different dimensions: measurement dimensions (pixel values, obtained directly from the original image), classification results (e.g. vegetation type, land use type, forested/deforested, crop type), instantaneous values of measurable phenomena or indicators. Moving up conceptually, this could be seen as a series of geofields (Câmara, Thomás et al. 1999) that overlap in space materialized as arrays. In order to make the semantic dimensions more evident, a mapping towards implementation as a set of arrays, to fit SciDB's physical model, would be needed.

Regarding the theoretical aspects, the dichotomy between of the pixel origin, strongly related to the measurement, and its application, aggregated in highly semantic concepts needs to be further studied and developed. We used two concepts, object and objectives to understand this dichotomy and later on propose a way to link pixels to higher semantic concepts.

With regards to implementation much needs to be done. With the Semantic Pixel, we intend users to be able to browse the information contained in the images in two different dimensions, using its semantics and its measurement characteristics. Users with different skill levels will use such a system in different ways, choosing to start with the measurement characteristics or the semantic values to perform their information retrieval, according to their individual preferences.

Acknowledgments

Authors acknowledge the support of CNPq (308678/2012-5, 401822/2013-3), and FAPEMIG (CEX-PPM-00518/13), Brazilian agencies in charge of fostering research initiatives.

References

Câmara, G., U. Freitas, et al. (1994). "A model to cultivate objects and manipulate fields." Proceedings 2nd ACM Workshop on Advances in GIS: 20-28.

Câmara, G., R. Thomás, et al. (1999). Interoperability in Practice: Problems in Semantic conversion from current technology to OpenGIS. Interoperating Geographic Information Systems, Springer: 129-138.

Couclelis, H. (1992). People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. A. U. Frank, I. Campari and U. Formentini. New York, Springer-Verlag. 639: 65-77.

Ferreira, K. R., G. Câmara, et al. (2014). "An algebra for spatiotemporal data: From observations to events." Transactions in GIS 18(2): 253-269.

Feyerabend, P. K. (1993). Against method. London; New York, Verso.

Fonseca, F. and J. Martin (2007). "Learning the Differences Between Ontologies and Conceptual Schemas Through Ontology-Driven Information Systems." JAIS - Journal of the Association for Information Systems - Special Issue on Ontologies in the Context of IS 8(2): 129-142.

Fonseca, F. T., C. A. Davis Jr., et al. (2009). "Spatial Data Infrastructures for the Amazon: a First Step towards a Global Forest Information System." Earth Science Informatics 2(4): 188-191.

Kant, I. (1996). Critique of pure reason. Indianapolis, Ind., Hackett Pub. Co.

Koenderink, J. J. (2005). Geometric Framework for Image Processing. Handbook of Geometric Computing. E. B. Corrochano. Berlin, Springer: 171-202.

Kuhn, T. S. (1996). The structure of scientific revolutions. Chicago, IL, University of Chicago Press.

Kuhn, W. (2009). A functional ontology of observation and measurement. International Conference on GeoSpatial Semantics - GeoS, Berlin, Springer.

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. Criticism and the growth of knowledge. I. Lakatos and A. Musgrave. Cambridge, England, University Press.

Paul, G. B. Overview of sciDB: large scale array storage, processing and analysis. Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. Indianapolis, Indiana, USA, ACM.

Popper, K. R. (2002). The Logic of Scientific Discovery. London; New York, Routledge.

Sinton, D. (1977). The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study. First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Dedham, MA, Laboratory for Computer Graphics and Spatial Analysis.

Stonebraker, M. (2010). "SQL databases v. NoSQL databases." Communications of the ACM 53(4): 10-11.

Weinberg, J. R. (1960). An Examination of Logical Positivism. Paterson, N.J., Littlefield Adams.