Volunteered Geographic Information and GIScience

Werner Kuhn
Vespucci Initiative and
Institute for Geoinformatics
University of Muenster
kuhn@uni-muenster.de

Position Paper for the NCGIA and Vespucci Workshop on
Volunteered Geographic Information; Santa Barbara, CA, December 13-14, 2007

Abstract

What are the research questions posed by Volunteered Geographic Information (VGI)? Does Geographic Information Science (GIScience) appropriately address them or does it need to shift its attention? I discuss changes to current research agendas, propose an arising grand challenge, and outline some specific new research challenges for scientists working within and outside GIScience.

Introduction

In early 1999, having just returned from a workshop which assessed basic research needs in GIScience [Mark 2000], I read the following sentence in a piece of the Economist on electricity markets [Economist 1999]:

> "Eventually, today's huge power stations and national transmission grids might be superseded by a system that relies on efficient local "micropower" generators."

It occurred to me then that the "national transmission grids" of mapping agencies and other oligopolies of geographic information (GI) might one day be superseded by efficient local "microGI" generators. I considered writing a column on this idea, but having to write regular columns didn't appeal to me then and I dumped the analogy as too far from reality. Here we are, a decade later, scratching our heads what the "microGI" or VGI generators out there mean for GIScience and GI markets.

Like for GIScience in general, there are two roles for VGI in science: scientific questions posed by the phenomenon of VGI, and the use of VGI in doing science. Research needs to integrate the two, because GI applications motivate and ground information science research, and better information theories improve applications [Mark 2000]. The main impact of VGI will probably be on how science and society work with GI: spatial has finally become normal in many respects, though it remains special in others. This evolution cannot fail to profoundly affect a majority of natural, technical, and social sciences, not just GIScience. Evidence is already provided by the knowledge (or cyber) infrastructures evolving in many fields, often with some spatial backbones.

One way to look at VGI is as the human side of the sensor revolution [Goodchild 2007]. In fact, the rapid expansion of sensor-based GI and that of VGI do seem to go hand in hand and share many characteristics, such as the distribution of information sources and their dynamic and low production cost nature. Also, the two developments individually and together allow significantly more powerful analyses and predictions, benefitting from broader observation coverage and better statistical filtering possibilities. They represent the "input" side of the rapid convergence of spatial technologies with information and communication technologies and communities, and in particular of the internet with the real world [Economist 2007].

GIScience started as "the science behind the systems" [Goodchild 1992], so we should again ask about the scientific questions posed by the technological phenomenon of VGI. Is there really much novel about VGI for GIScience? I strongly believe there is. A quantum leap has occurred in our ways of "spatial data handling", changing the practice and science around GI dramatically. It can be characterized as the *scaling up of closed loops*. Closed loops are systems incorporating feedback. An example of a closed loop is your controlling of water temperature in the shower. In this paper, I claim that

1. GIScience needs to develop theories and methods to control the daily "shower" of spatially referenced data; and
2. VGI is the "hot water".

Andrew Frank has suggested the notion of "closed loop semantics" [Frank 2007] to capture how feedback loops in information systems ground the *semantics* of information. This role in grounding meaning is a core aspect of VGI, as I will discuss below, but closed loops are not just improving semantics. They impact the entire breadth of the science behind spatial data handling. To combine the shower and power metaphors, closing the loop through VGI advances us from an age where one had to go to public baths for personal hygiene to one where we largely control our own flow of water.

The remainder of the paper surveys possible changes to existing research agendas, outlines the overall new research challenge for GIScience, and derives some specific research questions from that.

Old research challenges revisited

From 1995 onwards, the University Consortium for Geographic Information Science (UCGIS) has regularly identified research challenges, culminating in the set of 20 long and short term research priorities of 2004, elaborated on in [McMaster and Usery 2004]. The eleven long term priorities on this list are:

1. Spatial Ontologies
2. Geographic Representation
3. Spatial Data Acquisition and Integration
4. Remotely Acquired Data and Information in GIScience
5. Scale
6. Spatial Cognition
7. Space and Space/Time Analysis and Modeling
8. Uncertainty in Geographic Information
9. Visualization
10. GIS and Society
11. Geographic Information Engineering (distributed computing, SDI, data mining)

Focusing on grand challenges for GIScience, the 1999 workshop cited above identified the following four issues (slightly rephrased here):

1. *representing* the infinitely complex world in limited computing systems;
2. characterizing the *differences* between digital representations and reality;
3. improving transitions between *cognitive* and computational representations;

4. making *simulations* of geographic phenomena more realistic.

Additionally, the workshop report pointed out the "data challenge", i.e. the need for

5. coping with the *increasing quantity* of data being collected and archived.

A year later, the Association for Geographic Information Laboratories in Europe (AGILE) produced a Green Paper on its own Research Agenda, outlining the following five challenges (also slightly rephrased) [Craglia et al. 2001]:

1. understanding the *social aspects and policies* of geographic information;
2. constructing a comprehensive theory of spatio-temporal *information management* and presentation;
3. producing *dynamic models* of environmental and social processes;
4. achieving *semantic interoperability* for spatial data and services;
5. *bridging the conceptual gaps* of how space and time are viewed from different disciplines.

To determine how VGI affects these challenges, one can sort them into three categories: those advancing in their significance, those holding up, and those declining. Such a classification is obviously partial and I only present my guess. I have removed duplicates from the above lists, rephrased again, and attempted some rationalization for the class assignments:

Advancing

The advancing issues, i.e., those becoming more important through VGI, can again be sub-classed into those pushed by the mere fact of having a supply of "hot water" (VGI), those addressing the use of it, and the opportunities created by it.

"hot water" supply

An increase by orders of magnitude in providers of data (those volunteering GI) means growing needs to manage, filter, and integrate data reflecting different world views. This has already advanced research needs in the areas of

- Spatial Ontology
- Spatial Information Acquisition and Integration
- Spatial Cognition
- Geographic Information Engineering
- Theory of spatio-temporal information management and presentation.

For example, many semantic web researchers are addressing needs at the intersection of semantics and web2.0 technologies; navigation data providers have caught up on exploiting their users as data sources, but lack theories and methods for filtering and integration; search and archiving companies like Google and others are trying to get their hands on whatever information they can, but have to rely on not much more than unstructured tags (if any) for discovery and integration of contributed resources; cognitive differences in conceptualizing, locating, and expressing volunteered information typically outsource integration to those accessing the information; information engineering for distributed computing and information infrastructures has barely begun to take VGI seriously; and all these developments reveal that our theories of information management and presentation are mostly stuck in a traditional mind set of static databases where consistency is the main goal and redundancy should be avoided.

Their inability to provide mappings between multiple and conflicting conceptualizations has now become a major bottleneck.

“hot water” use

Putting the hot water to good use requires more research on

- Space and Space/Time Analysis and Modeling
- Uncertainty in Geographic Information
- GIS and Society

With the vastly increased, often near real-time availability of spatially referenced information, analysis capabilities grow significantly. While it has often been said that availability of data is not a bottleneck to better analysis (but discovery, interoperability, and better models are), VGI does produce a qualitative change in availability. The likeliness of getting rapid access to data from “human sensors” in a certain region has become much higher. At the same time, the reliability of having continued access to them may not be guaranteed. This creates a need for a more dynamic configuration of analysis models, making them adaptable to appearing and disappearing coverage in observations. It also creates new and exciting analysis capabilities that belong into the opportunities section below.

A similar qualitative shift occurs in our research on uncertainty. So far, many vagueness and uncertainty issues resulted from the fact that spatial information, if available at all, was often coming from a single source. With VGI, the law of big numbers kicks in. This alters many uncertainty issues radically. For example, accuracy and reliability of road network information looks quite different when the data stem from a student collecting data every few years for a company or when thousands of drivers do it implicitly and explicitly every day.

The biggest change of emphasis in research directions due to VGI use, however, is occurring in the area of GI and society. Suddenly, the picture is changing from expensive GI trickling down to citizens from governments and industry, to GI generated by citizens and potentially useful for governments and industry. Policies of GI may now have to address privacy and liability issues much more than pricing and access. The privacy and liability issues arising when correct or faulty information provided by somebody with good intentions but without qualifications, or by somebody ill-intended, is used in professional or otherwise critical tasks are substantial. To exaggerate just slightly: while the discussion on how much a citizen should pay for information that has been collected with her taxes continues, the citizen is already collecting the information herself or from her peers. The central policy issue seems to be a decision on how much the users should be involved in the production and maintenance of GI. This decision is not supported well by existing research on business models, economics, and legal frameworks.

“hot water” opportunities

VGI also improves the conditions for doing some types of research, and can therefore be expected to boost it, particularly

- Simulations of geographic phenomena
- Dynamic modelling of environmental and social processes.

While many simulations can run without actual observations, some of them depend strongly on initial values, boundary conditions, or calibrations, which could be VGI.

More importantly, VGI has obvious potential for providing in situ data for validation. For example, a transportation modeller can now get much better information on whether the traffic jams are where the traffic flow model predicted them.
What applies to simulations is even more important for modelling in general. Automated and human sensors have already dramatically improved abilities to model natural and social phenomena realistically and to evaluate these models. When ground truth is volunteered, a lot more opportunities arise for dynamic modelling and phenomena that are a lot more complex become amenable to it.

What effects VGI will have on

- grid computing

and associated research is not clear. One could argue that once GI is produced locally, it can be processed locally as well, at least for some tasks, without a need to "gridify" the processing explicitly. Computational grids may then start to resemble electricity power grids. On the other hand, grid architectures may represent key opportunities for localization of GI production and use in the first place.

Holding up

If we consider the modeling and storage side of representation to be dealt with by several challenges above, then the main part left for

- Geographic Representation

is visualization. This aspect of representation seems not particularly changed by VGI. Appropriate visualizations on a rapidly growing spectrum of displays remains an important research challenge, though, and solutions to some aspects of the problem (like visualizing point clouds, dynamic networks, or uncertainty) become particularly pressing facing VGI.
The understanding and handling of

- Scale

in GIScience remains unsatisfactory. Smooth transitions between representations at multiple scales, in models as well as displays, are still rare. Cartographic and model generalization are one aspect of them, but the reverse processes (densifying models) and the question how to select appropriate scales for modeling are key research challenges in the field. They may in fact grow in importance, due to the more dynamic, multi-faceted ways of representing GI, but probably not due to VGI per se.

Declining

It is not clear whether any current research topic will really decline in their importance due to VGI. If anything, one may argue that

- Remotely Acquired Data and Information,

hugely important as it is, provides the backdrop against which VGI is collected and geo-referenced, but does not create new research challenges in this combination. However, with the evolution of geobrowsers based on high resolution imagery, this verdict may look one day like that of Thomas Watson, chair of IBM, predicting in 1943 that there is a world market for maybe 5 computers.

New research challenges

Knowledge has been described as "actionable information", implying that it affords to use information in making decisions and carrying them out as actions. By closing the loop leading from reality through information to decisions and actions (figure 1), VGI can be seen as resulting from "informationable action", i.e., as information generated through actions in the world, actively or passively. For example, change requests from navigation system users to road databases result from driver observations and then support navigation decisions of other drivers, which in turn may provide their own change requests etc. Similarly, a driver may opt to make some of the floating car data generated by on board sensors available to others (e.g., on temperature and humidity). Or, in the scenario of figure 1, drivers can change their route based on traffic congestion information, thus reducing the congestion.

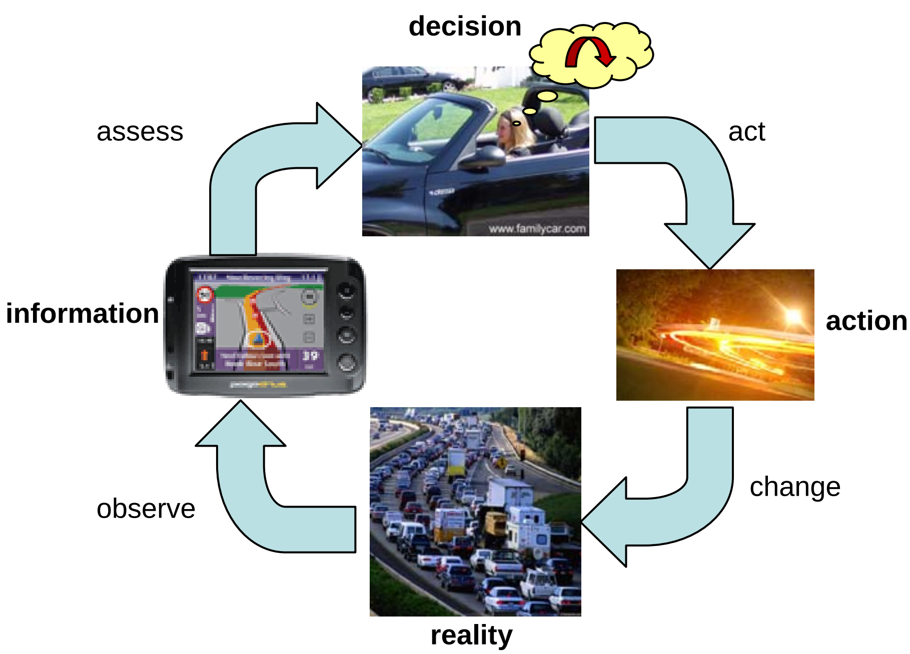

Figure 1: The information-action cycle

What does this idea of "closing the loop" imply for research needs and opportunities beyond traditional GIScience agendas? There are many possible ways to answer this question. I will look at research induced by VGI from four perspectives: technology, semantics, cognition, and society.

Technology

Let us first look at the engineering and technology research challenges posed by VGI. More of them will emerge once the current experimental and grass root supply of VGI has turned into more stable "hot water" infrastructures. Yet, some research needs and opportunities have clearly appeared already.

From an engineering perspective, one of the most exciting developments happening with VGI is the convergence of tools and processes for top-down and bottom-up modeling. The most active area may be the integration of semantic web and web 2.0 developments. The targeted result has been labeled “web 3.0”, even specifically acknowledging the key role of geospatial information in it [Economist 2007]. One of the main scientific issues here is indeed how information can be integrated through locational reference, at a scale and in a diversity that was never even considered in traditional GIS or SDI architectures. This is more than a quantitative change regarding the volume of information; it raises by itself a broad range of research questions, such as how to

1. characterize the quality of VGI;
2. annotate VGI with useful metadata;
3. discover pertinent VGI;
4. integrate VGI across multiple sources and with traditional information.

There is also a convergence of VGI with ongoing attempts to capture original (rather than derived) data. VGI creates a new kind of measurement-based systems [Buyong and Kuhn 1990, Goodchild 2004], supplying sensor-generated measurements with human observations.

The biggest technology push resulting from VGI, however, is a shift from conventional combinations of complex formats with simple data serving API’s (Application Programming Interfaces) to combinations of smaller (micro) formats with more versatile processing API’s. When there are relatively few data providers in a domain, contents can be standardized at the format level (using, for example, GML feature types), and exchanged over data access protocols (as in Web Feature Services). This is what OGC and ISO standards currently support. With the unbounded variety of VGI contents, however, standardizing their feature models may become impractical. What VGI needs to be shared and re-used are

5. simple formats to capture contents, and
6. flexible API’s to access and manipulate it.

This is a shift in emphasis, but an important one. Through the increased distribution of data and processing, it becomes possible to keep API’s slim and focused, while increasing their combined processing power. The success of API’s for VGI (such as the one of Google Maps) clearly shows a winning strategy: mash-ups supporting the ad hoc integration of heterogeneous contents through multiple lean API’s, instead of a few fat services and systems.

There is even a further reaching pay-off that we can expect from this standards evolution:

7. a bottom-up definition of geospatial processing services through API’s.

Given the difficulties to come up with a generic but powerful interface standard for OGC Web Processing Services, and the lacking development of feature operations as part of the OGC General Feature Model, a “grass roots” movement showing useful processing services is more than welcome and creates lots of engineering research opportunities.

Semantics

When GIS came from a few vendors and data models from few agencies and companies, the need for semantic modeling was limited. VGI, however, comes with vastly different conceptualizations underlying its large variety of sources. Accounting for these conceptualizations and mapping between them is an essential requirement for understanding and using some VGI at all.

While the spatial and temporal references in most VGI can easily be mapped across multiple reference systems, this is not (yet) the case for thematic references. For example, when bird watchers contribute bird sighting information, they have to choose a species standard. These standards differ in the way they assign species to birds, so that the same bird is classified differently in different parts of the world [Mark 2007]. Mappings are often easy to establish, but there are no tools and infrastructures in place where they can be defined and computed. While this example represents only a simple case of semantic mismatches, there is an unlimited multitude of them, but no theory that would help to classify and resolve them.

The biggest novel challenges in the semantics area, however, are to

8. exploit the grounding effect of VGI on semantics;
9. enable and capture semiosis in the social networks around VGI;
10. combine ontologies with folksonomies;

By the first challenge, I mean the above mentioned idea of "closed loop semantics", where actions in the world depend on the interpretation of terms in GI and thereby ground the semantics of the terms. For example, I learned the meaning of the US traffic sign "Do Not Pass" when I first saw it on a road where I could simply not stop and just had to move on… Some researchers in semantics believe that the grounding of meaning in action is the only solid foundation for the semantics of information. But, in order to exploit it, we need theories of the interplay between information and actions that are more specific and detailed than decision theories or ideas from action research.

The second challenge builds on the first and takes it further toward an understanding of how (technical) language evolves in information communities. Collaborative tagging constitutes a huge play ground on which "kids" start to apply certain terms to what their peers can see and these buy into their language use or override it with something else. This is how natural language evolves through pidgins. The process has huge potential for letting GI-related terms and their semantics evolve in dynamic user communities, rather than being dictated through "feature attribute catalogues". Today's collaborative tagging mechanisms, however, are only weakly supporting this evolution of tag semantics. For example, a participant at this meeting posted the tags he uses for two classes of VGI sites by Email to a discussion list. But this shows the need to support dynamic semiosis, and tag suggestions mechanisms (for example in del.ici.ous) are a good start. We need a deeper understanding of the semiotic processes occurring in technical language.

Such social or collaborative tag systems are called folksonomies. They obviously constitute a powerful resource for classifying and annotating GI. The third challenge stems from the fact that it remains unclear how (bottom-up) folksonomies are best integrated with (top-down) ontologies, especially during the genesis of both. How can ontology engineers be guided through evolving folksonomies, and how can a tagger be guided by pre-existing domain structure in some form of an ontology? Avoiding a quality degradation of ontologies in the first case, and an over-constraining of language production in the second are tough problems requiring research from engineering, linguistic, and social perspectives.

Cognition

Research on cognitive aspects of GI has been flourishing in three main areas: spatial cognition, cognitive engineering, and cognitive semantics. I address the first two here, having already dealt with semantics above.

Spatial cognition research has already been working from a "closed loop" perspective for a while. Research programs like that of the Transregional Collaborative Research Center "Spatial Cognition: Reasoning, Action, Interaction" [1], emphasize an integrative cognitive view of the cycle from acquisition through organization to utilization and revision of knowledge about spatial environments. As such, they are already studying the cognitive foundations for VGI and may not find major new research challenges from it.

Yet, some areas of investigation in spatial cognition are re-emerging. For example, "vernacular place" research. In the early nineties, some computer scientists became fascinated by the geographic idea of "place", as complementing the computational geometric spaces [Erickson 1993, Freksa and Barkowsky 1996]. Through VGI in the form of geo tags and the more visible role of individual and group conceptualizations of space through VGI, there is now a revival of research on vernacular place names.

Research in the area of cognitive engineering has typically looked at interface design for novel devices (handheld, public displays etc.). Important insights for VGI production and use conditions result from this. However, there is a lack of interaction research driven by the medium (the spatially enabled world-wide web) and user activity contexts (e.g., moving in space), rather than by new hardware. Single line query windows and a patch works of plug-ins and RSS feeds in our browsers do not scale to VGI infrastructures, whether they are on handhelds or on desktops. Like Smith and his co-workers at Xerox in the 1970s [Smith et al. 1982], we need to

> 11. find metaphors that help people interact with the web while communicating about their spatial environment.

Designing simplicity and elegance into interfaces going beyond text search, and combining them with innovative VGI generators (such as Google's image labeler game) are just two aspects of this challenge. They raise the same fundamental questions as thirty years ago: what is easy and what is hard for humans to do, but for novel applications (not office automation anymore) and for users engaged in social networks that are online (not outside the system).

In a broader view of cognitive aspects, one involving social psychology and cognition, many exciting research challenges emerge as well, though I don't feel competent to list and discuss them. One of the most often cited questions concerns how to

> 12. explain and exploit people's motivation to volunteer GI.

[1] http://www.sfbtr8.spatial-cognition.de/

This broader perspective leads to my last challenge area – that of the impact of VGI on social and institutional research. I feel equally incompetent to address it and will just highlight a few issues.

Society

Since this survey of research challenges focuses on information science questions, it cannot do justice to the vast area of social science research suggested by VGI. Yet, VGI accelerates a convergence of computational and social perspectives on information science, mainly in three areas:

13. model trust and reputation in online communities;
14. develop business models for producing and using VGI;
15. protect privacy and intellectual property.

Trust tops the famous semantic web layer cake. Its importance has dramatically increased with the arrival of VGI and other community generated information. The traditional assumption that only government agencies are trustworthy GI providers has rapidly collapsed (similarly to that of encyclopedia publishers regarding trustworthy knowledge sources). While the notion of trustworthy information still needs to be defined exactly (is it based on trust in people or trust in information, and how are the two related?), it is obviously tied to the reputation of information providers, as evidenced by online systems for auctioning and e-business. Traditional information quality parameters like accuracy, consistency, and completeness are rarely available or even meaningful in a VGI context. Since trust and reputation models have been useful in filtering other collaborative web content, one could use them as proxies for GI quality. But are existing trust and reputation models valid and useful for GI? What would suitable spatio-temporal extensions involve? A further reaching Darwinian analogy suggests that VGI could behave as memes, competing for adoption based on perceived trustworthiness and fitness for use.

Traditional business models for GI face similar challenges from VGI like those of the music industry. While research has narrowly focused on pricing of GI and failed to address value questions and maintenance models, entire sectors of GI have suddenly come under pressure by community-generated online maps, imagery, and services. This healthy development of closed loops around users as providers creates exciting opportunities for sound economic analyses and business strategies. It is a pity that few economists seem interested in the special nature of GI. Their advice is badly needed by GI providers, in order to decide what production and maintenance tasks they can "crowdsource", i.e. for what products they can tap into their customer base to provide, complement, or maintain it. Answering this question rapidly and intelligently will be critical for survival, but needs a much better understanding of GI as a commodity than we have it today.

Finally, many legal aspects of GI appear in a different light facing VGI. What if the new providers of GI care much less about intellectual property than traditional ones do? What if privacy becomes can be regulated by those affected directly rather than governments and enterprises? Research on these and other legal aspects is becoming more important than ever, but in a setting that involves new players and new attitudes.

Somewhat ironically, the difficulty of solving legal (as well as economic) questions pragmatically for traditional GI has impeded its markets and accelerated the VGI movement. Sadly, the hands of innovative mapping agencies appear to be tied by legal frameworks enforcing competition rules that

These brief spot lights on some social aspects of a VGI research agenda may only serve to make one point: that VGI is really more of a social than a technical phenomenon. Consequently, VGI can be expected to expand the attention of the interdisciplinary field of GIScience from a focus on engineering and humanities to the social sciences.

Conclusions

This overview of some research challenges changed and posed by VGI is necessarily incomplete and partial. I adopted a technology push view of research in this area, because the whole field of GIScience has been driven by such pushes. GIS itself came about when planners and surveyors wondered what the computer could do for them and went on to revolutionize geography, surveying and related fields. Now, we are wondering what neo-geographers can do for the world. As "spatially aware *non*-professionals", they are the new kids on the block of spatial data handling, looking perhaps more attractive than they will eventually turn out to be, but certainly raising lots of interesting questions for science. The 15 new challenges I identified above are exciting enough to me to claim that, in closing the loop of GI, we have found a grand challenges for GIScience.

References

Buyong, T. and Kuhn, W. 1990. Local Adjustment For A Measurement-Based Multipurpose Cadastre Systems. XIX Congress of FIG. Helsinki, Finland: FIG, vol. 3: 19-27.

Craglia, M. Gould, M., Kuhn, W., Toppen, F. 2001. The AGILE Research Agenda. 7th EC-GIS Workshop. Potsdam, Germany: European Commission. http://www.ec-gis.org/Workshops/7ec-gis/papers/html/agile/agile.htm

Economist, 1999. Power to the people. 18 February 1999. http://www.economist.com/business/displaystory.cfm?story_id=E1_TRSDPQ

Economist, 2007. The World on your Desktop. 6 September 2007. http://www.economist.com/science/tq/displaystory.cfm?story_id=9719045

Erickson, T. 1993. From Interface to Interplace: The Spatial Environment as a Medium for Interaction. Conference on Spatial Information Theory, COSIT 1993.

Frank, A.U. 2007. Data Quality Ontology: An Ontology for Imperfect Knowledge. Spatial Information Theory, 8th International Conference, COSIT 2007. Melbourne, Australia: Springer LNCS 4736: 406-420

Freksa C. and Barkowsky, T. 1996. On the relation between spatial concepts and geographic objects, in Geographic objects with indeterminate boundaries, A.U. Frank and P. Burrough, Editors. Taylor and Francis: 109-121.

Goodchild, M.F. 1992. Geographical information science. International Journal of Geographical Information Systems. 6(1): 31–45.

Goodchild, M.F. 2004. A general framework for error analysis in measurement-based GIS. Journal of Geographical Systems. 6(4): 323–324.

Goodchild, M.F. 2007. Citizens as Sensors: The World of Volunteered Geography, in VGI Specialist Meeting Position Papers. 2007: Santa Barbara, CA. http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf

Mark, D.M. 2000. Geographic Information Science: Critical Issues in an Emerging Cross-Disciplinary Research Domain. URISA Journal. 12(1): 45-54. http://www.urisa.org/files/Markvol12no1-4.pdf

Mark, D.M. 2007. Personal Communication on Bird Species.

McMaster, R.B. and Usery, E.L. 2004. A Research Agenda for Geographic Information Science. CRC Press.

Smith, D.C.S., Irby, C., Kimball, R., Verplank, B. and Harslam, E. 1982. Designing the Star User Interface, BYTE **7**(4) 242-282. http://www.guidebookgallery.org/articles/designingthestaruserinterface